



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"**

**CORSO DI LAUREA MAGISTRALE IN
ECONOMICS AND FINANCE**

TESI DI LAUREA

**"FROM EX-POST TO EX-ANTE EVALUATION:
ESTIMATING RETURNS TO HIGHER EDUCATION IN THE U.K."**

RELATORE:

CH.MO PROF. ENRICO RETTORE

LAUREANDA: ANNA DINGILLO

MATRICOLA N. 1084304

ANNO ACCADEMICO 2015 – 2016

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.

Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

ABSTRACT

The dissertation I present investigates recent developments achieved in the field of policy evaluation, exploring the techniques that allow to perform some kind of analysis that look beyond traditional objectives: this is the essence of going from ex-post to ex-ante evaluation. Conducting an ex-ante evaluation of a public policy works toward exploring potential results of a reform never implemented. Actually, the power of ex-ante evaluation from a policy-maker perspective is great, since it allows to answer to a wide variety of policy questions. This work focuses on the evaluation of public policy concerning education, defining the theoretical framework developed over the last decades until the most recent findings. Specifically, I consider the education sector of the United Kingdom and perform an empirical evaluation of the returns to higher education in this country.

SOMMARIO

La dissertazione che presento studia gli sviluppi recenti realizzatisi nel campo della valutazione delle politiche pubbliche, esplorando le tecniche che permettono di intraprendere delle analisi che vadano oltre al raggiungimento degli obiettivi tradizionali: in questo si definisce l'essenza del passare dalla valutazione ex-post alla valutazione ex-ante. Condurre la valutazione ex-ante di una politica pubblica mira ad esplorare gli effetti potenziali di una riforma mai adottata. Effettivamente, la capacità della valutazione ex-ante dal punto di vista del policy maker è considerevole, dato che permette di rispondere ad una vasta varietà di questioni di politica pubblica. Questo lavoro si concentra sulla valutazione delle politiche rivolte al settore dell'educazione, definendo il contesto teorico sviluppatosi negli ultimi decenni fino ai risultati più recenti. Nello specifico, prenderò in considerazione il settore dell'educazione in Gran Bretagna ed eseguirò una valutazione empirica dei rendimenti dell'istruzione superiore in questo Paese.

INDEX

INTRODUCTION	1
1 FROM THE EX-POST TO THE EX-ANTE POLICY EVALUATION	3
1.1 Policy evaluation: definition and problems	3
1.2 A comparison of ex-ante and ex-post methods: literature review	6
1.3 Combining ex-post and ex-ante techniques, an example. (Education choices in Mexico: using a structural model and a randomized experiment to evaluate PROGRESA)	22
2 RETURNS TO EDUCATION	26
2.1 Homogeneous and heterogeneous treatment parameters	27
2.1.1 Estimating average effects	27
2.1.2 Estimating marginal effects	28
2.2 Estimating marginal returns to education: an example from the literature	30
2.2.1 The model	32
2.2.2 The Local Instrumental Variable estimator (LIV)	34
2.2.3 Policy Relevant Treatment Effects and Marginal Policy Relevant Treatment Effects ..	37
2.3 Estimation's procedure of MTE and Marginal Policy Effects	40
2.3.1 Sample Selection Model	41
2.3.2 Endogenous Treatment Effect Model	45
2.3.3 Parametric estimation	46
2.3.3.1 The two step estimator	46
2.3.3.2 The Maximum Likelihood estimator	47
2.3.4 The LIV estimator	49
3 FUNDING HIGHER EDUCATION IN THE U.K.	51
3.1 Higher education reforms in the U.K.: from 1960s to our days.	51
3.1.1 1960s – 1990s	52
3.1.2 1990s – 2000s	54
3.1.3 2001 – 2015	55
3.2 College participation: the effects of the reforms and other influencing factors	57
4 ESTIMATING RETURNS TO HIGHER EDUCATION IN THE U.K.	61
4.1 Sample selection	61
4.2 Dependent and Independent variables	62
4.3 Model's Equations	66
4.4 Estimation results	67
4.4.1 First-stage regression	68

4.5	Possible extension and the ex-ante evaluation of returns to education	75
CONCLUSIONS		78
REFERENCES		82

INTRODUCTION

The dissertation I present investigates recent developments achieved in the field of policy evaluation, exploring the techniques that allow to perform some kind of analysis that look beyond traditional objectives: this is the essence of going from ex-post to ex-ante evaluation.

Conventional ex-post evaluation relies on consolidated methods (e.g. instrumental variables, difference in difference, regression discontinuity design, etc.) that attempt to extract a causal relationship from data with the highest grade of internal validity, ensuring at the same time the generalization of results to the external environment, thus the external validity.

Clearly, this is the first, important and most of the time arduous step to complete. Indeed, any other further grade of analysis cannot be achieved if the ex-post evaluation does not accomplish these purposes. This consideration is the key point for performing successfully ex-ante evaluation.

Conducting an ex-ante evaluation of a public policy works toward exploring potential results of a reform never implemented. In particular, we can ask ourselves: given the existence of a certain policy and its causal effect, what would be the set of policy rules that allows to achieve a bigger effect? Actually, the power of ex-ante evaluation from a policy-maker prospective is great, since it allows to answer to a wide variety of policy questions. It is obviously interesting to understand and forecast what would be the most effective means (e.g. a subsidy, a training program, the construction of a new school, etc.) that permit to maximize a certain measure of output (earnings, employment rate, volume of participants, level of personal skills, e.g.) or alternatively, understand if it exists a different set of policy tools that allows to achieve the same result while respecting the Government budget constraint. With the suitable validated model, ex-ante evaluation permits to fulfil useful cost-benefit analysis.

Feeling confident of having the availability of the right model is the key point explained above. As discussed in *Chapter 1.*, having the possibility of relying on a natural experiment would be the perfect starting point to perform all the steps that leads to the ex-ante evaluation. In this case, validating the model chosen for the ex-ante analysis means applying it to the data provided by the experiment to verify if it permits to replicate the results of the ex-post one. The validation process and also the choice of the suitable model, has a particular importance and various techniques can be followed. *Chapter 1.* provides a review of the literature that

goes through all these aspects and many others, comparing the result achieved by different authors.

This work focuses on the evaluation of public policy concerning education, defining the theoretical framework developed in last decades until the most recent findings. Specifically, I consider the education sector of the United Kingdom and perform an empirical evaluation of the returns to higher education in this country.

The education sector has addressed the attention of researchers for decades given its importance for the economy of all countries. Reforms intended to promote education at all levels has to be developed carefully since human capital accumulation is one of the main drivers of economics growth. For this reason, Governments should always intervene with the appropriate measure to handle this phenomenon.

When conducting analysis concerning educational output, the main obstacle is represented by the unobservable individual characteristics that are considered a source of bias in the evaluation of the returns to educational qualification. The relevant literature has progressed around the management of consequences implied by these unobservables factors, discussing which are the most effective methods to be used and the most important parameters of interest to be evaluated. *Chapter 2.* considers these aspects, with a focus on the literature that starting from the 2000s, has been concentrated on exploiting all the methodologies developed and results achieved over years, to provide the elements that allows to perform ex-post and ex-ante evaluation in the field of educational policies.

In *Chapter 3.*, the features of the higher education sector of the U.K. are presented, with a description of the main reforms undertaken by the Governments over last fifty years in order to funding a sector that has experienced a huge increase in the number of participants over time.

Finally, *Chapter 4.*, illustrates the empirical results obtained from the application of the model presented in previous chapters, discussing limits, difficulties and suggesting possible extension.

1 FROM THE EX-POST TO THE EX-ANTE POLICY EVALUATION

1.1 POLICY EVALUATION: DEFINITION AND PROBLEMS

Heckman and Vytlačil (2000, 2-3) define the policy evaluation problem as “the problem of comparing outcomes of a policy in place with outcomes under alternative policies” and the problem of causal inference “consists of determining which causes affect outcomes and measuring their quantitative importance”. They continue specifying that: “The policy evaluation problem is a special case of the problem of causal inference which entails comparisons between a hypothetical state and the observed state where the “causes” are different policies”.

The *causal effect* is defined by authors as the change of outcomes for an agent across states (s, s') under the *ceteris paribus* clause stated by Marshall (1890), that means that only states (s, s') are varied.

Subsequently, Heckman and Vytlačil (2007a, 4790-2) indicate three main policy evaluation problems:

1. “Evaluating the *impacts* of historical interventions on outcomes including their impact in terms of welfare”.
2. “Forecasting the *impacts* (constructing counterfactual states) of interventions implemented in one environment in other environments, including their impacts in terms of welfare”.
3. “Forecasting the *impacts* of interventions (constructing counterfactual states associated with interventions) never historically experienced to various environments, including their impacts in terms of welfare”.

The first policy evaluation problem consists in that of *internal validity*: identifying one or more given treatment parameters in a given environment. The second one consists in the problem of *external validity*, defined as the evaluation of one or a set of treatment parameters estimated in one environment to another environment; the term environment refers to different groups of people or different time periods with respect to those that are object of

study, but sharing the same characteristics. The last policy evaluation problem consists essentially in *the ex-ante policy evaluation*.

Authors define the term *impact* as the construction of individual level or population level counterfactuals and their valuations. Evaluating impacts of historical interventions on welfare means conducting a welfare evaluations, either ex-ante or ex-post, of the outcomes derived from agents and/or society behavior and interventions.

With respect to the last two policy evaluation problems, authors specify that “these forecasting problems are special cases of the problem of causal inference in which extrapolation from knowledge of currently experienced states is required to forecast and evaluate states not previously experienced” (see Heckman and Vytlačil 2000, 3).

What I am going to present in the following chapters involve precisely the achievement of the three aspects of policy evaluation. The path begins by exploiting ex-post methodologies, both *internally* and *externally* valid, in order to develop a credible approach that is able to make ex-ante policy evaluations. As Di Nardo and Lee (2011) say, the main aim of ex-post evaluation strategy is to achieve the highest degree of internal validity, “a high degree of confidence that what is measured indeed represents a causal phenomenon”. They state that these methodologies pursue a goal which is complementary to that of ex-ante policy evaluations. Indeed, “(...) we view “external validity” to be the central issue in an attempt to use the results of an ex post evaluation for an ex ante program evaluation”. (see Di Nardo and Lee, 2011, 13). The main difference between the two approaches is that in the ex-post evaluation the credibility of the analysis depends on the “credibility of the statistical model of the experiment” whereas, in an ex-ante evaluations the credibility is focused on the “statistical model of the *behaviour* of individuals”, and the aspect that make the second one more hard to achieve than the first is that, in ex-ante analysis, validation occurs in context which are different than those in which data have been collected.

Attaining this objective implies unavoidably the need of managing some conventional estimation problems known in the policy evaluation literature. By adopting the notation used by Heckman and Vytlačil (2007a), you can interpret $Y(s, \omega)$ as the outcome corresponding to the state (policy or treatment) s for agent ω , with $\omega \in \Omega$, realized after the treatment’s choice. Before knowing the treatment, agents can make forecasts about it. In fact, it is exactly the

potential influence derived from these forecasts about future outcomes that can generate the problem of *selection bias*.

An individual when considering whether to participate or not can be influenced by his/her considerations about futures potential outcomes and this give rise to a selection problem. The greatest advantage of social experiment is precisely that of avoiding self-selection problems that could give raise to selection bias. When the treatment assignment rule is randomization, receipt of treatment is independent of outcomes of treatment.

Problems arise when assignment depends on choices of agents. An example is given by the *Roy model*, in which agents choose which treatment has to be received by themselves or other agents (e.g. parents choosing for their children) by evaluating potential returns of alternative treatments. In a “utility-maximization framework” agents choose the one that provides the highest income, and this self-selection (as I will show in more details later) does not allow to estimate properly the outcome (see Carneiro, Heckman and Vytlačil, 2010).

Continuing with the same notation, authors define the *individual level causal effect* for individual ω that compares objective outcomes of treatment s with those of treatment s' as: $Y(s, \omega) - Y(s', \omega)$, with $s \neq s'$.

Clearly, it is not possible to observe the same individuals in both states of the world; this problem is known as the *fundamental problem of causal inference* (see Holland, 1986).

This is a central problem in policy analysis and typically, it is handle by trying to estimate a population version of the individual level parameter defined above, therefore a *population level treatment parameter*.

Indeed, Heckman and Vytlačil (2007a, 4802) state that: “The conventional approach in the treatment effect literature is to reformulate the parameter of interest to be some summary measure of the population distribution of treatment effects like a mean or the distribution itself rather than attempting to identify individual treatment effects”. I will turn to this point later when discussing the parameters of interest in policy evaluation.

The second way to handle the *fundamental problem of causal inference* is recurring to the *structural econometric analysis*. Under a well specified economic theory it is possible to model $Y(s, \omega)$ in all its determinants. Thus, it is possible to understand the mechanism generating outcomes and choices of agents modelling also their dependence.

“Constructing this counterfactual in a convincing way is a key ingredient of any serious evaluation method”. As pointed out by Blundell and Dias (2009), it exists three main classes of policy evaluation methods: the experimental method that exploits randomized experiments, the non-experimental methods and the structural methods.

The *assignment rule*, so the way in which individuals are allocated to one group or another in the program or the way in which they receive the policy, is pure randomly in the social experiment, while in non-experimental method the researcher tries to mimic the randomization exploiting non-experimental data.

1.2 A COMPARISON OF EX-ANTE AND EX-POST METHODS: LITERATURE REVIEW

The aim of this section is to investigate the recent developments in the evaluation of public policies. One of the main concern of this field is trying to implement an efficient method to evaluate public interventions in order to evolve from the so-called ex-post approach typically followed in the treatment effect literature. What I mean is, finding a way that allows the researcher to design an as less costly as possible evaluation structure but at the same time achieving a degree of credibility of the tool as high as possible. Technically this methodologies fall within the field of the ex-ante evaluation processes and they are implemented through structural models. Many researchers in last years have studied this topic moved by the need of evolving from the experimental approach. Indeed, a controlled experiment is often not possible to implement, can be too costly and time-consuming and the consequently ex-post analysis may highlight the need of further adjustments of the policy or some elements of it; but again this is costly and time-consuming.

These difficulties may be overcome, at least partially, by the introduction of ex-ante methodologies. These ones anyway, are not free of disadvantages, as I will explain later. However, exploiting the advantages of both approaches may lead to an efficient and credible evaluation.

Heckman (2010) in his work introduces how to merge ex-ante and ex-post evaluation. He illustrates the pros and cons of both “structural” and “policy evaluation” approaches.

This distinction finds its roots in the previous distinction of the approaches pursued in evaluating microeconomics problems; (see Heckman and Vytlačil, 2000). Authors compare

the “structural approach” with the “treatment effect approach”; they argue that a structural model is designed to achieve different objectives, such as providing a framework for causal inference, evaluating the effects of different policies in place and performing forecasting analysis to construct counterfactuals of current policies or of policies never implemented. The treatment effect approach instead pursue a more precise objective: policy evaluation. Forecasting is not included in this approach.

Heckman (2010) underlines the fact that one important difference between *structural approach* and conventional *program evaluation approach* (treatment effect approach) is that the first makes explicit assumptions about the behavior and decision-making process of the agents whereas the latter does not. The program evaluation approach just emphasizes the power of the randomized experiment that allows to identify the outcome without having to explain how preferences of agents are formed, how the mechanism determining counterfactual states acts or which are the sources of variability among agents. Therefore, structural approach focuses on the *causal mechanism*, program evaluation approach on the *causal effect*. The first allows forecasting the effects of policies which have never been implemented, the second just analyzes the effects only after the policy’s implementation. This distinction however, can be seen not just as a limitation of one of these approaches or of both of them, but instead as the key that can be used as a leverage to make one method as the continuation of the other.

Pronzato (2012) in her work makes a comparison between the quasi-experimental approach and the structural one by showing their implementation in the analysis of a reform of lone parental welfare. She works on the same data and uses the same outcome variable for the two strategies. For the quasi-experimental evaluation, she considers the sample of Norwegian lone mothers for the treatment group, and mothers in a couple for the control group, both before and after the reform, estimating the effect with the triple-difference method.

The sample for the structural model consists of a sub sample of just lone mothers observed before the reform, a construction chosen for a typical ex-ante evaluation, as we will see in other works. Here the message she gives is: “the two strategies help the understanding of policy impact in a complementary way: while the focus of the quasi-experimental evaluation design is to measure what really happened, the challenge of the structural model is to predict what potentially can happen”. (see Pronzato, 2012, 17-8)

This study provides results of the two methodologies which are very close. This can suggest that the reconciliation between the two improve their credibility. Indeed, when the predictions obtained by the structural estimation match the results of the quasi-experimental analysis one can say that the model is validated. But note that the validation applies mutually: it makes both methods credible. When this happens, the researcher can feel safer that the predictions about future policy's development is more robust; moreover, this allow to limit the cost of policy implementation.

The complementarity of the two approaches appears even more clear when highlighting the limitation of using just one of them. For instance, in her work the author specifies the fact that in the quasi-experimental approach we can observe how mother's behaviour change after the implementation of the reform, but we cannot disentangle different effects of the different parts of the reform. Consequently, it is not even possible to predict the most suitable policy's improvement that could be implemented in the future unless another reform is undertaken. But, implementing a structural model full specified in its parameters, permit to overcome this obstacle.

Clearly, even the structural approach presents some drawbacks. In particular, its computational complexity leads to achieve replication and sensitivity analysis not so easily. Model parametrization can be most of the time very difficult. Indeed, the model has to rely on a certain number of assumptions concerning functional forms and distributions of unobservables. Anyway, even in this case connecting ex-post and ex-ante analysis simplifies the work.

In their work, Duflo, Hanna and Ryan (2012) use both a randomized experiment and a structural dynamic model of labour supply to test whether monitoring activity and a set of incentives increase teachers' presence at school in India. Why have they chosen this technique? Because making just an ex-post analysis prevents from the analysis of the effects of some alternative kinds of incentive schemes that differ from the one experimented. Moreover, since they put teachers in the treatment group under control (by monitoring daily their presence at school) in addition to grant them with a non-linear incentives' scheme, the use of the structural model allows to disentangle the effect of the monitoring from that of the financial incentive. Also in this case the structural model is estimated using just the daily attendance data in the treated schools. This procedure is called *holdout samples* and is useful for the validation of the model. Here, in particular, the treatment group is chosen for the

estimation since the financial scheme provides the necessary variation for model identification.

To test the sensitivity of the model they construct different specifications of it based on different kinds of assumptions concerning unobserved heterogeneity and the error structure.

As authors say: “A primary benefit of estimating a structural model of behaviour is the ability to calculate outcomes under economic environments not observed in the data”. (See Duflo, Hanna and Ryan 2012, 1265-6).

Indeed, just thanks to the structural model they are able to identify the cost-minimizing combination of the elements of the policy: the amount of the incentive and the minimum number of days of work that each teacher has to complete to get it.

Maybe one of the most known work in the literature scene is the one of Todd and Wolpin (2010a). They exploited the randomized social experiment conducted by the Mexican government in rural areas consisting in a conditional cash transfer program called PROGRESA, to assess its effect on children’s school participation. Since the randomized experiment, even if simple to estimate, avoids the evaluation of alternative designs of the program, they adopted a behavioural model to estimate the effect of never implemented programs; they made therefore an ex-ante evaluation.

In their previous work (see Todd and Wolpin, 2006) they implemented a discrete choice dynamic programming model where parameters are estimated by simulated maximum likelihood; here, they exploited child wages variation across untreated villages, without using the variability induced by PROGRESA. Experimental variation is used by the author only for validate the model. In the subsequent work (see Todd and Wolpin, 2010) they investigate the use of a non-parametric dynamic model, that means a model where assumptions about functional forms are not specified. This aspect represented a great achievement in the attempt of reducing the computational burden of structural model but it is not free from drawbacks.

In general, behavioural models are applied to programs that affect the budget constraint, in particular modifying the costs side. When modelling the behaviour of agents, the researcher does not need all those data required to implement a matching or control function approach about treated and untreated; this is why they can be applied for ex-ante evaluation. Authors specify that not even strong functional form assumptions are necessary: they exploit the

condition for which non-parametric policy evaluation are met for a variety of policy intervention.

This methodology really simplifies the computational complexity but at the cost of maintaining some strong independence assumptions on the distribution of observed heterogeneity.

Typically, for their identification, structural model need that data about the policy instrument provide a source exogenous variation. The policy variables in question, must affect only the budget constraint without affecting directly the outcome equation.

As authors say “Ex ante evaluation requires extrapolating from past experience to learn about effects of hypothetical programs”. (see Todd and Wolpin, 2010a, 262).

Todd and Wolpin perform nonparametric estimation of a cash transfer to parents conditional on their children attending school. This is a subsidy that should incentivize school attendance by influencing the family’s budget constraint. Since authors consider an initial situation in which school is free, they do not have past data about tuitions from which to extrapolate variation. So they can estimate the model even if there is not a direct variation in the data related to the policy instrument, because the cash transfer is considered as a wage subsidy that enters in the budget constraint. By studying wage variation in the data they can analyze ex-ante the policy effect; the comparison is made constructing two budget constraint that refer to individuals that differ only in wage level. The subsidy act only through the budget constraint so it represents the exogenous source of wage’s variation that permit to identify the model. This is a key assumption of this method.

Here, a sort of “matching” is performed by equating particular *functions* of observables, not the observables directly as in the traditional matching estimator used in the ex-post analysis.

The second key assumption is that unobserved heterogeneity is independent of wage and other variables entering the budget constraint. This assumption is very restrictive because preferences of individual that affect school participation are likely to be correlated with factors related to the conditioning set. Authors manage this problem by conditioning also on some observable individual characteristics, even if this step can be non-trivial and goes counter-current with respect to the goal of reducing the model’s complexity.

Anyway, the main two limitations of this approach are:

- The estimation is based on a comparison that match groups of individuals exactly just at two different points of the wage distribution;
- moreover, this method cannot be applied to all classes of programs.

Also in this article, we have seen an example of the holdout sample method. Indeed, authors say that they “match” untreated individuals with other untreated individuals. So, they use the control group for their estimation, which result is compared with that of the already realized randomized experiment that serves as benchmark to validate the behavioural model.

Todd and Wolpin (2010b) and Keane, Todd and Wolpin (2011), examines in depths the most important characteristics of the discrete choice static/dynamic programming models, widely applied to the ex-ante evaluation and that falls in the category of structural estimation.

Todd and Wolpin (2010b, 22) provide a useful definition of the discrete dynamic programming model, a model “in which agent make choices sequentially over time from a discrete set of alternatives as new information arrives to maximize their expected utility over some time horizon”. They review some studies concerning empirical policy evaluation of different policies in developing countries using this kind of model. Moreover, they introduce these applications with a precise analytical presentation of the features of static and dynamic programming models.

Discrete choice models, both static and dynamic, are based on the latent variable specification. This framework considers that individuals make a decision at discrete time intervals, choosing between two (or more) alternatives given different state of the world. The latent variable function determines the outcome of the decision process, given the difference in payoff from choosing one of the alternatives. The payoff is a result of a typical problem of maximizing the utility under the budget constraint. Therefore, the latent variable may entails considerations about revenues and costs, the history of past decisions and some observed and unobserved variables that influence the final decision (either contemporaneous or lagged and contemporaneous depending on whether the model is static or dynamic respectively).

The aim of structural model so far cited, is exactly estimating the parameters of the latent variable function, which is not observed. Todd and Wolpin (2010b, 23) define the structural estimation as “the recovery of fundamental parameters of behavioural models, such as utility or technological parameters”. The non-structural approach instead estimates only a certain function of the structural elements.

A paper which follows a methodology very close to the one of Todd and Wolpin (2010a) is that of Ranjeeta (2010) which studies the effects of a conditional cash transfers to improve education and health in Nicaragua; it is another example of validating a structural model with the results obtained by a randomized social experiment realized in 2000. Ex-ante evaluation consists in a semi-parametric single index model that once validated is applied to simulate two alternative policy scenarios. The author extends the approach followed by Ichimura and Taber (2000) and the one of Todd and Wolpin (2010a) we have seen, that consists in estimating reduced form equations using minimal assumptions of functional form and estimating semi-parametrically the behavioural model to compare the predictions of it with the result of the experiment.

In order to evaluate the structural model, he does not estimate the structural parameters but instead relies on exogenous variation in observed policy variables. Policy variables that affect only the budget constraint and not directly the outcome equation are school costs and full income.

School costs do not include tuition because school is free in Nicaragua; so they include exogenous expenditure faced by families independently of tuition. They are observed only for families that have already enrolled their children at school; in order to achieve observability for the whole sample of children, school costs are predicted. Survey used to retrieve the data contains information about full wealth of families.

Also in this paper, author matches untreated individuals with other untreated individuals and the group of the treated in the ex-post analysis is used to make comparison of results and validate the model. Ex-ante results perfectly predicts results from experimental evaluation.

Even in this paper the author underlines the sense of this methodology: “Comparing the ex-ante results to the experiment provides a way of validating the model used. The validated models are then used to simulate alternate policy scenarios” (see Ranjeeta 2010, 23).

The effects of alternative policy scenarios are estimated for two health outcome variables, health check of children below 3 years and full coverage of vaccination, and enrolment rate as school outcome variable.

Keane, Todd and Wolpin (2011) defines four possible approaches to estimate a dynamic model:

1. non-parametric, non-structural;

2. parametric, non-structural;
3. non-parametric-structural;
4. parametric, structural.

First of all, all approaches require an exclusion restriction. Recalling what we have seen in the paper of Todd and Wolpin (2010a), there was needed an exogenous source of variation of wages, provided in that case by the subsidy. It acts only through the budget constraint and does not affect directly preferences entering in the alternative-specific utility function. So wage variation is exogenous, is independent of preferences. This is the key assumption to identify the model. More generally, the exogenous variation provides a policy-relevant variation, necessary when estimating the effect of a policy. The effect of wage on participating decision is isomorphic to that of a subsidy. Having a policy-relevant variation allows to estimate the policy effect even without having direct variation in the policy instrument (like for example the variation in tuition). Both tuition and exogenous wage variation are independent of individual preferences affecting utility, so studying data about one of them allows to identify the model.

Typically, once the model has been identified, only the parametric-structural approach allows for counterfactual policy analysis. Todd and Wolpin (2010a) have nevertheless, made a non-parametric structural estimation of the conditional cash transfer program on the probability of participation of children to school. However, they demonstrate that this estimation approach is only feasible when wages offer are observed. Without observing wage for those who do not work more assumptions about functional forms are needed.

As I have already said, in order to achieve counterfactual analysis implementing structural-parametric estimation of static/dynamic programming models, we need functional forms and distributional assumptions. In particular, assumptions are needed about unobserved heterogeneity. As Todd and Wolpin (2010b, 29) say, it consists in “permanent differences across agents that potentially affect the decision that they make but that are unobserved by the researcher”. Assuming that the stochastic component of the model are mutually serially uncorrelated simplifies greatly the computational burden and the estimation procedure. Anyway, estimate the model with difference specification can be useful for its validation.

Todd and Wolpin (2010b) list the two most important reasons for which structural estimation is useful. The first is that, it allows to separate the effects of individual preferences

and opportunities (policy interventions) on final outcome, helping understand which aspects have to be modified for achieve a certain desired outcome. The second is that structural estimation allows to analyse policy interventions never implemented and that maybe cannot be implemented (e.g. for economic reasons) and therefore that could never be assessed. Moreover, typically the experimental design can assess only policy intervention of a limited duration; long term evaluation can be instead achieved with structural models.

After having described some basic features of structural models, I continue hereinafter, presenting some other empirical studies that combine ex-post and ex-ante techniques together.

Di Porto, Elia and Tealdi (2013) present an ex-ante and ex-post social program evaluation on labour tax evasion in Italy. So, they investigate how the combination of different policy instruments impacts the reduction of tax evasion without raising unemployment. In their ex-post analysis they study if the reform approved in Italy in 2003, concerning the legislation about temporary contract and apprenticeship, had some effect on the reduction of informal work. Then, “using the ex-post findings as a background to design a theoretical model”, they make an ex-ante analysis through a structural model to simulate different policy interventions (that include changes in tax burden, penalty fee for tax evaders, firing costs and type of contracts) in order to find the optimal one that allows to achieve the outcome desired. The ex-post analysis is achieved by a DiD and triple difference analysis. Then, they realize a continuous time search and matching model to evaluate (ex-ante) the effect of temporary contracts in the informal sector. The estimation is made by modelling the labour market both before the reforms when only permanent contracts were in place and post reforms when temporary contracts have been added (see Di Porto, Elia and Tealdi, 2013, 11-23). In this paper, parameters are calibrated choosing their values according to some sources: literature about the topic, Italian legislation and statistics provided by the National Institute for Statistics (ISTAT).

The findings of this research show that by just analysing the effects of the 2003 reform, temporary contracts alone are not an efficient instrument to drive the “emersion” phenomenon. As we have already see, ex-post evaluation alone does not allow to study the effects of the potential different elements of a reform combined together to find the optimal mix. Applying different policy mix through an experimental design would be prohibitively costly. Structural estimation overcomes this limit and, as authors did, allow to define an optimal policy mix. What maybe miss in this work is a validation mechanism. The structural

model calibration results somewhat complicated in particular when calibrating firing costs, as authors explain. (See Di Porto, Elia and Tealdi, 2013, 23-4). They do not implement the estimation via an holdout sample techniques in which they could have estimate the parameters modelling the labour market just before the reform and then comparing results with the post-reforms analysis.

Wasmer (2012) evaluate the 1989 welfare policy reform implemented in France. It consists in studying the effects on employment of a living allowance granted to all individual satisfying certain requirements. The interesting technical aspect of this work is the fact that the author calibrates a matching model with the estimates provided by the ex-post analysis realized with a DiD method. It can be noted that, contrary to the calibration realized in the work mentioned before (see Di Porto, Elia and Tealdi, 2013) here, Wasmer applies this method in order to calibrate and validate at the same time its model. He first implements a difference-in-difference and triple difference-in-difference method identifying the control and treatment groups that allow to control for different regional trends and also performs various robustness check and falsification exercises. Once he feels safe that the estimates are robust, he uses them to calibrate the key parameters of the labour market model to run a number of counterfactual policies including also the most recent French reform (of 2007) that provided changes in the previous one implemented in 1989. The steps he follows are the following:

- estimate the coefficient of the model prior to the reform to achieve some targets observed in the market pre-reform;
- calibrate the model using the DiD estimates of the economy post reform;
- finally, once the model is fully parametrized run counterfactual experiments (e.g. see the employment effects that would have been obtained in 1989 if 2007 reform had been implemented at that time).

Wasmer (2011, 30) says: “our results are a first step toward integrating ex-post estimations of public policies into ex-ante structural approaches”.

Geyer, Haan, Wrohlich (2012), estimate an intertemporal structural model of labour supply for mothers with young children receiving some governmental benefits and then validate their model exploiting a parental leave reform introduced in Germany, to define a natural experiment.

The structural estimation is made by modelling the market under the reform and taking data from the German Socio-Economic Panel Study (SOEP) in order to simulate the effect of the reform of 2007 and then, the same reform is analysed through a natural experiment using a different source of data. Results are compared and model validated. The validation of the model provides encouraging results suggesting that the structural model can be used to estimate the causal effect of a policy reform.

Even this work provides an evidence that combining experimental and structural approach can avoid many evaluations problems; indeed, authors say that (see Geyer, Haan, Wrohlich 2012, 1) “it is often criticized that structural models rely on strong assumptions that need to be imposed. Therefore, opponents of the structural approach question the reliability of those policy evaluations and instead suggest to exploit true exogenous variation for the identification of the causal effect on behavior induced by a policy reform”.

Brewer et al. (2006) have studied the impacts of a change in in-work benefits (the Working Families’ Tax Credit – WFTC) on labour market behaviour of families with children introduced in the U.K. in 1999. They designed a discrete choice structural model using micro-data before and after the transfer program to evaluate the program’s incentives of the participation rate (through participation costs) and the effects on labour supply.

In this work, the advantages coming from the structural model highlighted by the authors are:

- separating effects of in-work benefits from other contemporaneous taxes and benefits changes;
- controlling the program’s self-selection effects.

The program was implemented at a national level and participation was contingent on eligibility criteria. Therefore it has been not design as a controlled experiment and this imply the lack of a control group.

Authors indeed, model a structural model of labour supply including participation of eligible individuals; they follow a non-parametric identification approach relying just on functional form assumptions and considering as the source of variation the changes in taxes and benefits over time and different eligibility status of individuals, both acting through the budget constraint.

Having the data available, they validated the capability of the model to capture the labour force participation, hours worked and program participation by comparing the predicted values with the actual ones. They subsequently perform some counterfactual analysis concerning the effects of alternative combination of the tax and benefit system both for lone mothers and couple families. Finally, they compare their estimations with those of other studies concerning ex-ante evaluation of the same program with a structural model, ex-post evaluation exploiting a natural experiment and estimation of the same program in other countries.

This work represents an example of another possibility to design a structural model to evaluate a counterfactual policy starting from the actual state of the world that supplies the necessary data to design the framework and to test the estimation of the suggested model. Moreover, without having a controlled experiment that provide the basis for the comparison, it is still possible to try to validate the model gleaned from other studies and analysis both ex-ante and ex-post.

Blundell (2006, 424-5) says: “As a precursor to the analysis I will have to convince you of the validity of the structural model estimates. For this I will make a comparison with a simple difference in difference evaluation strategy. Although not providing sufficient information for policy simulation or the assessment of optimality, simple difference in difference evaluations can be valuable for validating the specification of more fragile microeconomic models”.

Remaining in the field of tax credit policies reforms in the U.K., Blundell (2006) evaluates the optimality of Earned Income Tax Credit Policies for lone parents in the U.K. and compare tax credit policy reforms in the U.S. (EITC) and the U.K. (WFTC).

His methodology consists first in assessing the validity of a dynamic model of labour supply by comparing the estimation of the impact of the implemented policy with that of a difference in difference analysis. Once the structural model is validated he studies the optimality of tax credit policies. The analysis of optimality is driven by the consideration that many times labour economics just focuses on the analysis of the average impact of a reform. Here the author wants to investigate whether the mentioned policy is optimal for low income individual, this means focusing of the intensive margin¹ of labour supply responses. The

¹ As Blundell, Bozio and Laroque (2013, 2) define: “(...) we split the overall level of work activity into the number of individuals in work and the intensity of work supplied by those in work. This reflects the distinction between whether to work and how much to work at the individual level and is referred to, respectively, as the extensive

objective is maximizing a well-behaved social welfare function subject to a government budget constraint.

Even in this example, author notices that: “On their own quasi-experimental approaches do not identify all the parameters necessary to assess optimality” Blundell (2006, 433). This is because in their analysis they need to have an estimation of the elasticities of labour supply response, whereas quasi-experimental and experimental approaches estimate just the average treatment effect. Again, this represents the main limitation of these procedures that can be overcome through a structural approach. Nevertheless, authors themselves underline the need of exploiting in any case the contribution of this approach: “(...) they can be used to assess the validity of structural estimates of the elasticity parameters”, Blundell (2006, 433).

Indeed, author proceeds by investigating the literature to define the characteristics of the structural model of labour supply including take up rate, and adopting a matching difference in difference approach to define the policy impact using data before and after the reform comparing potentially eligible parents with not eligible one in the control group and making assumptions on unobservables.

The source of variation needed to identify the model come from housing costs and local taxation that, as always, acts through the budget constraints across individuals in the sample; instead, the specification of assumptions needed, in particular concerning unobserved heterogeneity, comes from related works.

Finally, he runs the simulation of WFTC policy reform. He studies how the labour supply behaviour of individuals varies when parameters of tax and transfer system vary.

Thoresen and Vattø (2013), put all their efforts in demonstrating how it is possible to reconcile the quasi-experimental approach with the structural one, given the growing dominance of the latter in the policy analysis. In the field of tax and benefits reform, authors follow the reasoning of Blundell (2006) and validate a discrete choice model of labour supply with a reduced form panel data analysis. An important topic in the field of taxation consists in the concept of elasticity of taxable income (ETI): a parameter that measures the response in taxable income to a change in the net-of-tax rate. They compare the “ETI literature” with the structural estimation. In the ETI literature, typically, panel data of actual labour income levels

and intensive margin of labour supply. At the aggregate level the former is typically measured by the number of individuals in paid employment and the later by the average number of working hours”

before and after the reform are used. The reform is necessary for identification, provided by panel data that gives net-of-tax rate variation across individuals and time. The approach involves the individual's utility maximization subject to a budget constraint, using IV techniques to deal with endogeneity problems, and employing the difference-in-difference estimator.

As authors say: “although the discrete choice labor supply model continues to be a key instrument for predicting policy changes, serious concerns have been raised about the ability of structural models to generate robust predictions about the effect of policy changes, (...), it is essential to use other source of information to validate the models” (see Thoresen and Vattø, 2013, 5).

In this peculiar side of the literature, the comparison necessary for the validation of the structural model has to be made carefully since the ETI methodology estimates the average treatment effect for the treated, while the labour supply model gives us responses that differ along the income scale. To do so, first, the structural model estimate earning pre- and post-reform under exogenous wage assumption and finally, the same regression framework of the ETI approach is used to estimate elasticities for the simulated earning levels. These estimation are compared with estimation obtained using reduced form panel data analysis. Results are similar.

The aim is the one common for all policy analysis: the ETI methodology measures the average elasticities that follow a specific tax change reform and is not informative about other potential reform. Having a validated structural model may, therefore, overcome this limitation.

To remain in the sphere of taxation implications, Bourguignon and Ferreira (2003), review ex-ante evaluation techniques based on estimation and simulation of structural econometric models of household behaviour. They model labour supply effects of tax-benefits systems in developed countries and simulate effects of potential reforms in those countries. They underline how ex-post analysis are undoubtedly useful, even if they are necessary but not sufficient alone to allow the policy maker to define the proper policies to achieve some desired results. When an ex-post analysis of an *existing* policy shows that some features have to be reformed, it then will turn to be essential to determine a list of possible alternatives and find the optimal one in term of outcomes and implementation costs. In order to do so, authors

define some counterfactuals which depend on changing of household behaviour once reforms are implemented.

They affirm: “such an analysis is *marginal* because it is meant to capture differences from the status quo. Also, it is almost necessarily *behavioral*, because of the need to generate counterfactuals that take agent responses into account (...) this requires some *model*, which transforms the actual sample into the counterfactual one.” (see Bourguignon and Ferreira, 2003, 3).

Bernal and Keane (2010), compare the impact of maternal and alternative care providers’ time inputs on children’s cognitive development, since empirical studies in the literature show that this last one seems to be highly correlated with future labour market outcomes. In this interesting work, authors need to manage a problem of self-selection given that mother’s employment and childcare use decisions tend to be correlated with unobserved characteristics of mothers and children. Therefore, given a longitudinal data sample of lone mothers, they exploit some reforms concerning aids to families as an exogenous source of variation on the incentive to work/use childcare, essential to the identification of the effects of mother’s work/childcare decision on child outcomes. These effects are identified through a structural approach. Indeed, in their specification authors want to study the effect of the time dedicated by the mother to their children; for this reason, comparing just average outcomes under an instrumental variable approach may over-identify the impact since it will include the effect of changing not only time inputs but also good inputs. This aspect can be clearly handled by modelling a structural model of mother’s employment and childcare decision. Moreover, with such a model authors can manage the problem of unobserved heterogeneity affecting decision rules. In particular the methodology is called “quasi-structural” since they “approximate” decision rules for employment and childcare use and then estimates these ones jointly with a child cognitive ability production function and mother’s wage function. In this approach the difference with respect to a fully specified structural model is that the last one includes expectations about futures, in this case for example about changes in welfare rules.

As for all structural models, the identification is granted by a natural exclusion restriction which in this case is the set of reforms that affect the decision rules for employment and child care use but do not enter the cognitive ability production function. Additional instruments (local demand conditions) are adopted and act in the same way.

Also in this case, a natural experiment is involved as a tool of validation of the structural estimations. Authors find similar results between structural estimation and IV approach using same instruments even if the selection bias is handled differently in the second one.

It is interesting to note that authors list a series of previous works where problems of endogeneity have been handled differently, underlying how however, results seem inconclusive or that differ greatly between each other. They found that techniques that involve extensive sets of explanatory variables, fixed effects, value added models face with difficulty the problem of endogeneity or even not at all. Just the IV method deals well with it but, as authors say, using weak instrument ends up with a failure in estimation. To conclude, combining the power of using good instruments with a model that could deal with selection bias allow to obtain robust estimation.

When the literature provides several studies on a certain subject, it would be very useful to compare final results, model's formulation (a dynamic model rather than a static one, a binomial vs. a multinomial decision setting, and so on) assumptions (about functional forms, unobserved heterogeneity, individual preferences), to discuss estimation methodologies and obtain parameters estimates needed, for example, to calibrate the model.

Obviously, these synergies are useful in all kind of researches, but when performing ex-ante evaluation they are more. The goal is trying to develop a model investigating the actual context but also the economic theory to find the optimal estimation solution that suite the concrete specific case and is able to predict future scenarios. All this, having in mind the goal of replicability of the analysis and of the credibility of its predictive power. Modelling individuals behaviour in a predictive way can maybe be challenging and understand the optimal mix of behavioural assumptions can involve a great effort, as we have seen, that sometimes could reveal to be ineffective and imprecise.

Ex-post evaluations that rely on experimental or quasi-experimental design do not really have to face behavioural modelling problem; their concern is observing what is happened and choosing the most suitable technique to analyse data and handling unobserved ones with consolidated techniques.

The trade-off depends on the extent of the analysis. Ex-post analysis allow to achieve results that could be both internally and externally valid in a rather consolidated fashion, but it does not allow to explore beyond one specific question. Structural models instead, require a

greater effort in term of knowledges and computational burden in order to achieve some degree of replicability and validity, but with a notably greater analytical potential.

This is why resolving the trade-off can be achieved through both approaches and when studying public interventions' implications with structural models, it is very important to having a variety of examples as wide as possible available in order to resolve and simplify the drawbacks of the ex-ante techniques.

An intelligent combination of all the methodologies insofar saw could, maybe, smooth out criticisms about policy evaluation techniques. Extrapolating the best from each approach allows to define a new efficient one that can answer to a wide range of policy analysis' issues.

Heckman (2010, 2) comparing structural models and “treatment effects” literature say: “The two approaches have much to learn from each other. A more active dialogue would benefit practitioners of both approaches. (...) A synthesis of some of the best features of both approaches would produce a better approach to the evaluation of social policy and of medical procedures”.

1.3 COMBINING EX-POST AND EX-ANTE TECHNIQUES, AN EXAMPLE. (EDUCATION CHOICES IN MEXICO: USING A STRUCTURAL MODEL AND A RANDOMIZED EXPERIMENT TO EVALUATE PROGRESA)

Attanasio, Meghir and Santiago (2012) evaluate PROGRESA, evolved subsequently in *Oportunidades*, the same program implemented in Mexico and previously studied by Todd and Wolpin, (2010a). The aim of the paper is still the same: analyse the effects of monetary incentives to education choices in that country, in order to incentivize school enrolment of poor children. By exploiting the rich dataset provided by the randomized social experiment, authors show how to combine them efficiently with a structural model of education choices in order to make ex-ante analysis. The difference with respect to the work of Todd and Wolpin, (2010a) is that here Attanasio, Meghir and Santiago (2012) choose the way of dynamic structural parametric estimation. Contrary to Todd and Wolpin, (2010a) that did not exploit the variability induced by PROGRESA, Attanasio, Meghir and Santiago (2012) estimate a structural model identified by the variation induced by the experiment (such that it is for sure exogenous) using both the treatment and control group. They stress the point that the marginal utility of the grant is different from the marginal utility of other sources of income, like wage. Therefore, exploiting the variation of the opportunity cost of schooling (wage), like Todd and

Wolpin (2010a), may have different implications on the program's effect estimation. Moreover, authors estimate the general equilibrium effects that the program could have on children wages. They found that the program resulted in an increase in wage in the treatment group by reducing labour supply of children.

The randomized experiment's estimation suggest an effect of the program of the type of an inverted U-shaped, with a small impact for children aged 0-10, a peak between 10-14 and then declining again.

The model is designed as follows: each child faces two alternatives: schooling or working; one option precludes the other; going school envisages the grant provided by the program and costs affected by observable and unobservables factors. Children are allowed to go to school up to the age of 17, so at 18 they recover the investment, modelled by a terminal value function. The grant received by PROGRESA is compared with the monetary reward of going work.

The reason for choosing a dynamic model is motivated by the fact that when deciding about enrolment, each children face each year an option to continue or drop taking into account the structure of the program: grants are available for the last 3 years of primary school and the first 3 year of secondary school. Completing primary school gives eligibility for receiving secondary school's grant. Then, the terminal value function depends on the highest grade completed. So current decisions affect future ones; there is a state dependence because the number of years of school completed affects utility of attending the current one.

Authors also test for anticipation effects, given that in some areas the implementation of the program is delayed, but they have not found anyone.

A careful specification of all costs and benefits of attending school is essential: benefits depends on the utility of attending school, childcare services provided by the school and past attendance. Household entitled to PROGRESA living in treatment villages receive the grade and gender-specific grant.

Costs involve buying all what it is necessary to attend school, from books to clothes and transportation; in addition, it is considered the opportunity cost of attending school represented by the lost opportunity of receiving a salary.

So, utility function of attending school (u_{it}^S) depends on the grants received (g_{it}) and on the remaining pecuniary and non-pecuniary costs or gains from attending school (Y_{it}^S). (Y_{it}^S) equation includes: a vector of taste shifter variables (that exclude household income since it is likely to be endogenous); the variable denoting attending primary or secondary school and the related costs; an extreme value added term i.i.d. over time and individuals, an element that introduces dynamics (in the sense that current schooling choices affect future grades and utility costs); and finally, a term representing unobservables (for the econometrician, but known for the individual) assumed to have constant impact over time.

The utility function of not attending school (u_{it}^W) depends on the potential earnings when out of school (w_{it}) and a costs-benefits variable (Y_{it}^W) which is a function of the same factors above described, a part from school attendance and related costs.

The model includes uncertainty represented by: first, the future costs of schooling that since affect future schooling choices also affects the current ones and second, by the possibility of failure in completing a grade.

Value functions (for schooling and working) that takes the form of a Bellman equation, are modelled to design the comparison between current costs of schooling and future benefits and costs.

Wages represent the opportunity costs for education; since wages are not observed for children who do not work, and the dynamic programming requires that individual predict future wages, authors model a wage equation. The equation serves to test the presence of general equilibrium effects of the program and predictions from this equation are used in place of actual wages.

Finally, the presence of past education that creates dynamic effects involve an initial condition problem since the econometrician is not able to observe the entire history of schooling of each child. Authors solve the problem specifying a reduced form for educational attainment up to current date.

Exogenous variability in the data is derived from the fact that among treated villages families eligible are those classified as “poor”. Moreover, the grant varies by grade attended but in the same grade there could be children of different ages. Therefore, the effect of the grant is identified by comparing across treatment and control villages, eligible and ineligible households and different ages within and between grades.

The model is estimated by maximum likelihood; different versions of the model are estimated and results compared. First of all, authors find that the dynamic model fits the data well and the estimation results match those obtained in the experimental design. The program effectively increase enrolment rate between primary and secondary school but the impact on children of primary school age is not big. Authors ran also simulation to changing some elements of the policies; in particular, in order to improve school participation more resources should be offered to older children. Results suggest that given the same amount of resources spent, the impact on school enrolment decision of older children is larger. The conclusion they stated is that the program should change its structure according to different age levels.

Authors underline how it is important to have a fully specified structural model to understand all the mechanisms underlying agents' behaviour. Only by modelling so accurately every aspects it is possible to understand where it is necessary to intervene and how; simulations are then useful to understand which future policies would be optimal in term of costs sustained and benefits achieved.

The evidence is that only the development of such a precise structural model produce the emersion of several results and considerations that are useful in order to improve the reform's design and achieve more effective results.

2 RETURNS TO EDUCATION

Estimating returns to education represents an area of interest in labour economics that has always drawn the attention because of the importance since ever conferred to the role of human capital in economic growth².

Education indeed, is a component of the concept of human capital, therefore several studies have looked for a causal effects of years and quality of education on labour market outcomes, estimating the “returns to education”. Two important works that left an important trace are those by Mincer (1974) and Griliches (1977).

In his famous book, Mincer (1974) developed the *Mincerian wage regression*, the starting point still currently adopted in this field of analysis, to represent the existing positive relationship between earnings, education and experience. This gave raise to further discussions concerning several econometric issues as Griliches (1977) pointed out in his work. For instance, he addressed the problem of the “ability bias”, discussing the role of ability in the Mincerian equation and the way it should be properly measured and interpreted. Moreover, when treating the endogeneity of schooling decision, he underlined an important aspect that prevents the estimation of the true causal effect of education. Griliches (1977,13) explained that:

“Schooling is the result, at least in part, of optimizing behaviour by individuals and their families. This behaviour is based on some *anticipated* earnings function. To the extent that the “errors” (...) in the ex-post and ex-ante earning functions are correlated, they will be “transmitted” to the schooling equation and induce an additional correlation between schooling and these disturbances.”

This discussion paved the way to the concept of *heterogeneity* in returns to schooling. It is just since recent decades that econometricians estimating returns to education has focused, more than the past, on the problem of *heterogeneity in returns to education*.

Card (1999, 2001) dealt with the estimation of the causal relationship between education and earnings in the presence of heterogeneous returns to schooling; he presented a survey of the literature that applies IV using institutional changes in the education system and discuss the discrepancies between IV and OLS estimates. The evidence is that in a wide number of studies the IV estimates are much higher than OLS ones. One possible explanation advanced

²See Mincer (1981), Solow (1957), Barro (1992), Becker (1962, 2009), Romer (1990), Lucas (1988), Schultz (1961), Jorgenson and Griliches (1967), just to cite a few.

by Cards, relies on heterogeneity in returns. In the model he adopted, based on Becker (1967), individuals make their schooling choice comparing costs and benefits, maximizing the discounted present value of net educational earnings. Given that aptitudes and tastes towards schooling vary among individuals, also marginal costs and returns vary and the optimal level of schooling will be heterogeneous.

He explained these findings suggesting that the marginal returns to schooling in particular for those individuals coming from less-wealthy backgrounds, are higher than average marginal returns for the population. This is deducted by the fact that IV estimates marginal returns only for a sub-population and is not adapt to a population level estimation. Card (2001, 1156) said:

“Institutional features like compulsory schooling or the accessibility of schools are most likely to affect the schooling choices of individuals who would otherwise have relatively low schooling. If the main reason that these individuals have low schooling is because of higher-than-average *costs* of schooling, rather than because of lower-than-average *returns* to schooling, then "local average treatment effect" reasoning suggests that IV estimators based on compulsory schooling or school proximity will yield estimated returns to schooling above the average marginal return to schooling in the population, and potentially above the corresponding OLS estimates.”

This consideration can be treated as an important starting point to discuss the most recent literature concerning the estimation of returns to education.

2.1 HOMOGENEOUS AND HETEROGENEOUS TREATMENT PARAMETERS

Estimation of returns to education has put the focus onto managing the analysis of heterogeneous returns. Under this framework, conventional population level parameters and estimation methods could not be suitable; in this section, I introduce the development achieved in the literature.

2.1.1 Estimating average effects

The conventional population parameters estimated in the treatment literature (using the notation of Heckman and Vytlačil, 2007a) are:

- $ATE(j, k) = E(Y(j, \omega) - Y(k, \omega))$

Where the mean effect of moving from program, state or treatment j to program, state or treatment k for individual ω is estimated.

Another conventional parameter is the Average Effect of the Treatment on the Treated (TT):

- $TT(j, k) = E(Y(j, \omega) - Y(k, \omega) | D(j, \omega) = 1)$

Where $D(j, \omega)$, is the treatment status that assume value 1 if the individual receives the treatment, and 0 otherwise. It estimates the mean effect of moving from program, state or treatment j to program, state or treatment k for those individuals who get the treatment.

Finally, in the same way, the treatment on the untreated (TUT) is defined in contrast as follows:

- $TUT(j, k) = E(Y(j, \omega) - Y(k, \omega) | D(j, \omega) = 0).$

All these are “average effects”, valid at a population level. When average effects are constant across individuals or “homogeneous”, average and “marginal effects” are the same. In case of “heterogeneous treatment effects”, the opposite is true. It means that a certain program, policy or reform, does not allow all individuals to receive benefits of the same magnitude. This would be an interesting phenomenon to explore, since for example from a policy maker point of view, it would be useful to identify what portion of population would benefit of greater gains from an intervention than another one.

Following the same line of reasoning of Card (2001), Angrist (2004) said that *Instrumental Variable*, identifies a causal effect which is internally valid for those individuals whose treatment status was affected by the instrument chosen. To be externally valid, the assumption needed is the one of homogeneity. “Basic IV assumptions identify causal effects on ‘compliers’, defined as the subpopulation of treated individuals whose treatment status can be influenced by the instrument” (see Angrist, 2004, C53). This definition relates to the parameter introduced by Imbens and Angrist (1994), the *Local Average Treatment Effect*.

2.1.2 Estimating marginal effects

The literature concerning heterogeneous effects has been a focus of several studies; we can find at first the work of Heckman and Robb (1986) that considered a model of selection into training when the effects varies among individuals, and adopted alternative evaluation methods to estimate the effects of the treatment in case of non-random selection.

Björklund and Moffitt (1987) produced one of the first work concerning returns to education under heterogeneous treatment effects. “One of the largest literatures in empirical

labor economics concerns the estimation of the determinants of the effects of various individual choices on wages”. (See Björklund and Moffitt, 1987, 42). In the literature many studies involve models of self-selection where individuals having different characteristics, both observables and unobservables to the researchers, internalize in their decision-process the knowledge and expectations they have formed about the realization of future returns, representing the outcome of a given educational path. This phenomenon is called indeed, *selection on gains*. Since returns are not constant across population, individuals self-select into the alternative that satisfy their *own budget constraint*. This means that the above described average population treatment effects do not coincide with the marginal ones.

The work of Björklund and Moffitt (1987) introduced some concepts that have been widely studied in all subsequent works: Heckman and Honoré (1990) assessed the conformity of the *Roy model* to study a wide array of occupational choices. This model “(...) explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills”. (See Heckman and Honoré 1990, 1121). Heckman, Smith and Clements (1997) stressed the importance of the existence of heterogeneity of impacts among individuals as opposed to the traditional evaluation literature that considers distributional issues of policy impacts as irrelevant. Finally, Heckman and Vytlačil (1999, 2000, 2005, 2007a, 2007b) developed a rich literature that addresses and unifies all these aspects.

The main concept of this branch of the literature is resumed in this sentence (see Björklund and Moffitt, 1987, 42):

“The implications of heterogeneity in rewards -or heterogeneity in the rate of return- are many and interesting. First, we are able to estimate not only the average wage gain to the activity of those who are currently participating, but also the marginal wage gain of the individuals who are on the margin (...) Second, we show that (...) we can distinguish the wage gain from the welfare gain to the activity”.

Authors contributed also to the definition of the *Marginal Treatment Effect* (MTE) a parameter of interest that researchers want to estimate under heterogeneous treatment effects framework. It can be defined as the impact of the treatment on those individuals at the margin of indifference about participation.

Subsequently, Heckman (2001b, 107) said:

“An important distinction is the one between evaluation models where participation in the program being evaluated is based, at least in part, on unobserved idiosyncratic responses to treatment and models where participation is not based on unobserved idiosyncratic responses. This is the distinction between selection on unobservables and selection on observables. The validity of entire classes of evaluation estimators hinges on whether or not they allow agents to act on unobserved idiosyncratic responses”.

In next section, I present technical aspects of the methodologies used to estimate marginal treatment effects.

2.2 ESTIMATING MARGINAL RETURNS TO EDUCATION: AN EXAMPLE FROM THE LITERATURE

Carneiro, Heckman and Vytlačil (2011) estimates the *Marginal Treatment Effect* of college participation exploiting a local version of instrumental variable.

The framework is characterized by self-selection and idiosyncratic returns entering the decisional process. Thus, marginal and average returns to schooling are ex-post not the same. Anyway, it is possible to recover marginal returns at different points of the margin of indifference. This is achieved through a local version of instrumental variable (see Heckman and Vytlačil 1999, 2005, 2007b). The analysis allows to study how to obtain a marginal expansion in college attendance induced by a variation in the available instrument that not necessarily has to correspond to the variation induced by the policy under consideration.

Using data from the National Longitudinal Survey of Youth (NLSY) of white males in 1979 they show that returns to college vary across individuals that act knowing their idiosyncratic returns to education. The MTE (*Marginal Treatment Effect*) is central in the analysis because it allows to define all the conventional parameters of interest but also others that answer interesting policy questions, like the *Marginal Policy Relevant Treatment Effects*. The MTE is estimated in their work using a robust semiparametric selection model and continuous instruments. Comparing the return (identified by these parameters) of one year into college with results given by OLS and IV estimators, they find that both OLS and IV estimation of the Average Treatment Effects are upward biased. This finding just confirms the evidence resulting from the literature in this field.

Returns to schooling are conventionally measured in monetary term, thus the impact that a certain number of years of education has on earnings. This is what precisely estimates the well-known *Mincer Equation*, stated as follows:

$$Y_i = \alpha + \beta S_i + \varepsilon_i \quad (1)$$

Where, Y_i is the log wage for individual i , S_i is a dummy indicating university enrolment, β is the parameter of interest identifying the return to education, and ε_i is the residual.

Carneiro (2003), gives a precise definition of the causes of all problems of estimation in the Mincer equation, analysing the sources of heterogeneity. The first problem arising in this regression is the correlation between S_i and ε_i , the latter interpreted typically as unobserved ability. This is called the “selection in levels”, the conventional source of bias considered in the literature, typically called “ability bias”. It can be simply solved through conventional methods, like Instrumental Variables. Authors found that estimating the *Mincer Equation* through conventional methods demonstrate that returns, β , vary across individuals, thus β is random. It is interpreted as the percentage increase in earnings due to an additional year of schooling. When it is correlated with S_i it give raise to the *selection on returns* or *on gains*, interpreted as the condition of those individuals that decide to go to school because they are aware of the benefits they can retrieve from schooling, thus they exploit their knowledge about their own idiosyncratic returns. This means that the variables that determine returns to education in the outcome equation are correlated with the variables entering the selection process.

Analysis made at the margin of indifference allows to recover the marginal return of a policy that expands the individual probability of attending university, thus the return of a policy that induces students, that otherwise would not have enrolled, to enrol. With standard IV procedures instead, the risk of overestimating returns is ran, because identification encompasses returns also of those students that would have enrolled a priori. Indeed, under standard IV procedures people induced to participation by a change in the instrument could not be the same as those induced to participation by a policy change. Authors found that the two procedures give results that differ substantially, with the IV results overestimated (0.0951 for conventional IV and 0.0148 for estimated marginal returns). This because, marginal expansion of university participation attracts students with lower net returns than those of students already enrolled.

Carneiro, Heckman and Vytlačil (2011) estimate the MTE adopting different estimation procedures in order to benchmark one against the other. The normal selection model gives precise estimations and show that:

- MTE, evaluated at mean values of \mathbf{X} , is declining in Us^3 meaning that people with highest gross returns to schooling are more likely to enrol into college, and vice versa. This supports the evidence that individuals self-select into the sector where they have a comparative advantage.
- Range of variation of returns is (-15.6%; 28.8%) but heterogeneity can be even larger considering two aspects:
 - The MTE is the average gain measured at each quantile of the desire to go to college;
 - Accounting for variation in \mathbf{X} , the range expands from -31.56% to 51.02%.

The semiparametric approach results confirms these findings, even if MTE estimates presents larger standard errors.

2.2.1 The model

Starting from the standard Mincer equation (1), Heckman, Carneiro and Vytlačil (2011) apply the *generalized Roy model*⁴ where:

$$Y_1 = \mu_1(\mathbf{X}) + U_1 \text{ and } Y_0 = \mu_0(\mathbf{X}) + U_0 \quad (2)$$

are the potential log wages that the individual would face if he decides to attend (Y_1) or not (Y_0) college, where: $\mu_1 = E(Y_1|\mathbf{X} = \mathbf{x})$ and $\mu_0 = E(Y_0|\mathbf{X} = \mathbf{x})$. \mathbf{X} are observable individual characteristics.

Returns to schooling, or the individual treatment effect, are identified by

$$Y_1 - Y_0 = \beta = \mu_1(\mathbf{X}) - \mu_0(\mathbf{X}) + U_1 - U_0. \quad (3)$$

As pointed out by Carneiro, Heckman and Vytlačil (2001), when $U_1 - U_0 \neq 0$, β varies in the population.

The average treatment effect is defined as: $\bar{\beta}(x) = E(\beta|\mathbf{X} = \mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$.

Writing the outcome equation in potential outcome notation

³ Notation is explained in next section; \mathbf{X} is the vector of observable individual characteristics and Us corresponds to different quantiles of the unobserved component of the index of the desire to go to college.

⁴ See Heckman and Vytlačil (2007b).

$$Y = SY_1 + (1-S)Y_0 \quad (4)$$

And substituting equation (2) in (4), we obtain:

$$Y = \mu_0(\mathbf{X}) + S(\mu_1(\mathbf{X}) - \mu_0(\mathbf{X})) + U_0 + S(U_1 - U_0) \quad (5)$$

Where $\bar{\beta}(x) = \mu_1(x) - \mu_0(x)$, therefore,

$$Y = \mu_0(\mathbf{X}) + S\bar{\beta}(x) + U_0 + S(U_1 - U_0) \quad (6)$$

Since participation into the program is voluntary, it is necessary to define the decision rule. To do so, a standard latent variable discrete choice model is adopted, where the individual's net benefit of college attendance is given by:

$$I_s = \mu_s(\mathbf{Z}) - V \quad (7)$$

S is the dummy variable indicating college enrolment, \mathbf{Z} is a vector of observable variables and V unobservables to the econometrician. I is a latent endogenous variable and $S=1$ if $I_s \geq 0$ and $S=0$ otherwise.

Assumptions stated are:

- i. V is a continuous random variable with a strictly increasing distribution function F_V ; it may depends on (U_1, U_0) in a general way.
- ii. \mathbf{Z} may include some or all components of \mathbf{X} plus some variables excluded from it (the exclusion restriction).
- iii. (U_1, U_0, V) is statistically independent of \mathbf{Z} given \mathbf{X} .
- iv. (U_1, U_0, V) is statistically independent of \mathbf{X} .
- v. $\mu_s(\mathbf{Z})$ is a non degenerate random variable conditional on \mathbf{X} (existence of a valid instrument).
- vi. Y_1 and Y_0 have finite first moments (in order to define mean parameters).
- vii. $1 > \Pr(S = 1|X = x) > 0$ for every $x \in \text{Supp}(X)$, this is required in order to assure the existence of treated and controls.

Monotonicity in \mathbf{Z} is also required, in the sense that given certain values of an instrument, z and z' , $S(z) \geq S(z')$ for all individuals, or $S(z') \geq S(z)$ for all individuals, meaning that a certain instrument either induces individuals into schooling or not.

The probability of attending college (the propensity score) is defined like: $P(\mathbf{z}) \equiv \Pr(S = 1 | \mathbf{Z} = \mathbf{z}) = F_V(\mu_s(\mathbf{z}))$, (conditioning on \mathbf{X} left implicit).

Under these assumptions, Vytlačil (2002) establishes that the model presented is equivalent to the LATE model of Imbens and Angrist (1994) under *Index Sufficiency*, that involves that \mathbf{Z} enters the conditional expectation only through $P(\mathbf{z})$.

The *Marginal Treatment Effect (MTE)* is given by:

$$\text{MTE}(\mathbf{x}, u_s) \equiv E(\beta | \mathbf{X} = \mathbf{x}, U_s = u_s) \quad (8)$$

With $U_s = F_V(V)$, uniformly distributed, where different value of U_s corresponds to different quantiles of V . Therefore, if $S = 1$, $P(\mathbf{z}) \geq U_s$. Representing MTE over different values of U_s , permits to see how “(...) returns vary with different quantiles of the unobserved component of the index of the desire to go to college” (see Carneiro, Heckman and Vytlačil, 2011, 2756). If the MTE does not depend on u_s , marginal and average returns are ex-post the same, so the parameter identified would be $\bar{\beta}(x)$, the average treatment effects.

The MTE is the mean return to schooling for individuals with characteristics $\mathbf{X} = \mathbf{x}$ and $U_s = u_s$, so for an individual indifferent between going or not to college, whose $P(\mathbf{z}) = U_s$. Equation (8) tells us that MTE is “the expected treatment effect conditional on the unobservables that determines participation” (see Carneiro, Heckman and Vytlačil 2010, 379).

Carneiro, Heckman and Vytlačil 2001, establish that the MTE is the limit of LATE when it exists.

2.2.2 The Local Instrumental Variable estimator (LIV)

The MTE can be estimated through the method of *Local Instrumental Variable*, by differentiating $E(Y | \mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p)$ with respect to p^5 , where, (leave the conditioning on \mathbf{X} implicit):

⁵ The importance of having continuous instruments is evident when computing this derivative; continuous instruments involve a continuous P , required to do this calculation.

$$\begin{aligned}
E(Y|P(\mathbf{Z}) = p) &= \mu_0 + \bar{\beta}p + E(U_1 - U_0|S = 1, P = p)p \\
&= \mu_0 + \bar{\beta}p + E(U_1 - U_0|P \geq U_s, P = p)p \\
&= \mu_0 + \bar{\beta}p + \int_0^p E(U_1 - U_0|U_s = u_s)du_s
\end{aligned} \tag{8}$$

Where $E(U_1 - U_0|P > U_s, P = p)p$ is a control function involved in selection on unobservables analysis, and

$$\partial E(Y|P(\mathbf{Z}) = p)/\partial p = \bar{\beta} + E(U_1 - U_0|U_s = p) \tag{9}$$

Therefore,

$$\text{MTE}(\mathbf{x}, p) = \bar{\beta} + E(U_1 - U_0|U_s = p) \tag{10}$$

is the *Local Instrumental Variable Estimator* (LIV) of Heckman and Vytlačil (1999).

A simple graphical observation of this derivative allows to understand the magnitude of the heterogeneity. A flat derivative indicates that the heterogeneity has little impact in the evaluation, whereas nonlinearity indicates that there is selection on unobservables. Since typically standard IV impose linearity in P , IV would be a valid estimator only in absence of selection on unobservables.

Indeed, estimating equation (6) with standard IV poses some problems; an instrument is valid if it affects the endogenous regressor and is not correlated with the error term in the outcome equation. From (6)

$$Y = \mu_0(\mathbf{X}) + S\bar{\beta}(x) + U_0 + S(U_1 - U_0) \tag{6}$$

It is clear that:

$$\text{plim } \hat{\beta}_{IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, S)} = \bar{\beta} + \frac{\text{Cov}(Z, U_0)}{\text{Cov}(Z, S)} + \frac{\text{Cov}(Z, S(U_1 - U_0))}{\text{Cov}(Z, S)}$$

even if $\text{Cov}(Z, U_0) = 0$, $\text{Cov}(Z, S(U_1 - U_0)) = \text{Cov}(Z, U_1 - U_0|S = 1)P$, therefore the last term is not equal to zero because $U_1 - U_0$ is dependent on S .

Linear instrumental variable is typically used when the goal is that of estimating the average returns to schooling, β . Under selection in levels, this method allows to estimate the average returns of schooling, either if β is constant or a random variable, but not correlated with S .

Things change under heterogeneity, when β is not only a random variable, but is also correlated with S . Under these circumstances is not possible to identify a parameter that averages out the distribution of returns.

Carneiro (2003, 4) says “(...) I find that the average person going to college has a higher return from the marginal person who is indifferent between enrolling in college or not. This suggests that heterogeneity is important and needs to be accounted for in policy analysis”.

$\hat{\beta}_{IV}$, estimated using $P(\mathbf{Z})$ as an instrument can be derived from a weighted average of the MTE (see Carneiro, Heckman and Vytlacil, 2001). Moreover, all other parameters of interest can be retrieved in this fashion, as it is explained in next section.

Given the empirical support of $P(\mathbf{Z})$ (conditional on \mathbf{X}), it is possible through the MTE to estimate the return to schooling for an individual indifferent between enrolling or not at all points of the margin of indifference, identified by unobservable factors entering the net benefit function. In other words, it is possible to identify the returns of a particular individual, identified by the quantile of the unobserved component of the desire to go to college (U_s), induced to go to college by a marginal change in $P(\mathbf{Z})$. By aggregating all instruments in $P(\mathbf{Z})$, it is possible to enlarge the support over which MTE can be estimated.

This method avoids problems that arise when estimating LATE in case of multiple instruments; estimating LATE in case of selection on gains using multiple instruments is typically achieved through varying one instrument at a time; anyway, this procedure needs to account for the covariation of each instruments with the others. Adopting $P(\mathbf{Z})$, it is possible to identify the contribution of each instrument in tracing out various regions of MTE function.

From (10), you can understand that an individual at the margin is an individual for which the net benefit function is null; those with an high value of U_s have therefore also an high value of $P(\mathbf{Z})$, for which ($U_s = p$). In order to induce him or her participating, it is needed that instruments assume certain values that makes the p-score through a marginal increase higher than U_s . Instead, those individuals with low value of U_s , with $P(\mathbf{Z})$ higher than U_s , face a positive net benefit function, thus are already enrolled; for these individuals, marginal increase in $P(\mathbf{Z})$ is worthless.

Once the MTE is estimated, Carneiro, Heckman and Vytlacil (2011) can retrieve parameters like the average return to college in the population, the average return to college

for students enrolled and also the standard IV estimator, just by creating weighted averages of the MTE⁶.

The general formula that allows to identify all parameters of interest starting from the MTE is the following:

$$\text{Parameter } j = \int_0^1 MTE(x, u_s) \omega_j(x, u_s) du_s, \text{ given } x \quad (11)$$

2.2.3 Policy Relevant Treatment Effects and Marginal Policy Relevant Treatment Effects

Another parameter of interest is the *Policy Relevant Treatment Effect (PRTE)*.

Once the MTE is identified, the optimal policy that induces individuals at the margin of indifference to enrol can be identified. How? Authors choose a class of policies that modifies the probability of participation (p-score). An exclusion restriction is needed, indeed Carneiro (2003), sets the conditions needed to evaluate the policy relevant effect; the policy under investigation has to affect only the selection equation in order to rule out general equilibrium effects (this is the standard requirement stated in all structural models as already presented in *Chapter 1*); then, the policy has to operate only through one of the instruments selected, changing their values inside the support of the data. In the estimation of education's returns typically, policies under consideration satisfying these conditions can be the change in tuition or distance to college. For example, Carneiro (2003), estimates the effect of subsidizing tuition in a fixed amount considering individuals that decide to enroll only if the subsidy is in place. The way in which this is modelled is the same already seen in other structural choice models (see Todd and Wolpin, 2010). The effect of a subsidy is equal to the reduction of tuition by a fix amount, e.g. $Z - \alpha$. Other subsidy schemes can be estimated too, like proportional tax reduction.

The formula is derived as follows; (keeping conditioning on \mathbf{X} implicit in all that follows) starting from a *Baseline Policy*, the aim is calculating the mean effect for a person experiencing a change of a policy that influences its probability of participation into a certain program. Define S^* as the treatment status under the *Alternative Policy*, Y^* as the outcome under the *Alternative Policy*, then if $E(S) \neq E(S^*)$,

⁶See Heckman and Vytlacil (2005) Table IB

$$PRTE = \frac{E(Y^*|AlternativePolicy) - E(Y|BaselinePolicy)}{E(S^*|AlternativePolicy) - E(S|BaselinePolicy)} = \int_0^1 MTE(u_s) \omega_{PRTE}(u_s) du_s \quad (12)$$

where, $\omega_{PRTE}(u_s)$ are the policy weights that depend on the distribution of P^* and P , the p-score under the alternative and baseline policy respectively. Therefore, this parameter identifies the average return for an individual that decide to enroll because the policy is in place but that would not enroll otherwise.

The PRTE depends only on the distribution of P^* , which is the probability of participation after the policy change. PRTE function links the distribution of P^* to the individual-outcome's change.

The limitation of this method is that the PRTE can be difficult to identified because it requires that the support of $P(\mathbf{Z})$ is the full unit interval, and it is not always the case. Indeed, as explained in Carneiro, Heckman and Vytlacil (2010, 386), “(...) suppose that the largest estimated probability of attending college is strictly less than 1. For analysing a tuition subsidy policy, it is possible that the largest probability of attending college under a tuition subsidy will be greater than the largest probability of attending college without a tuition subsidy, so the support condition for identifying the corresponding PRTE parameter is violated”.

The evolution of the PRTE that requires only a weaker condition, is the marginal version of the PRTE: the *Marginal Policy Relevant Treatment Effect* (MPRTE), that identifies the *marginal* change from a baseline policy. It requires weaker assumptions because it only needs that the MTE is estimated within the support of the data, therefore the full unit support of $P(\mathbf{Z})$ is not required. MPRTE is derived placing positive weights on the MTE (\mathbf{x}, u_s) , for those values of u_s where the density of $P(\mathbf{Z})$ is positive. Thus, identifying MPRTE is still possible even if PRTE is not. Moreover, the MPRTE is a parameter useful in conducting cost-benefit analysis of marginal policy changes.

Carneiro, Heckman and Vytlacil (2010) specify all the conditions required to estimate the MPRTE. In particular, they say that “the essential requirement is availability of a continuous instrument”. This means that the necessary assumption is that \mathbf{Z} contains a continuous variable, thus a continuous instrument for the treatment status variable. Under this condition, marginal policy changes can be analyzed and marginal policy treatment effects evaluated. The whole analysis is still conducted conditional on \mathbf{X} .

Using sequences of PRTEs it is possible to define a marginal version of this parameter of interest. Following Carneiro, Heckman and Vytlačil (2010, 2011), the MP RTE can be derived as follows.

Start by considering a baseline policy, for which the baseline probability that $D = 1$ is $P_0 = P(\mathbf{Z})$.

Now take a sequence of policies indexed by a scalar variable α , with $\alpha=0$ the baseline policy. Then, P_α is the corresponding probability of schooling, whose associated cumulative distribution function is denoted by F_α .

Next, for each policy α , define the corresponding PRTE parameter. The corresponding MP RTE parameter is the limit of the sequence of PRTEs as α goes to zero. An example that relates to following chapters, is a policy that affects tuition fees. If the k th elements of \mathbf{Z} is college tuition, and the policy subsidizes college tuition for example with grants, by a certain amount α , you will have $Z_\alpha^k = Z^k + \alpha$ and $Z_\alpha^j = Z^j$ for $j \neq k$. Carneiro, Heckman and Vytlačil (2011) estimate the MP RTE for policies concerning marginal change in tuitions and marginal changes in P .

Therefore, MP RTE is derived as follows:

$$MP RTE(\{F_\alpha\}) = \lim_{\tau \rightarrow 0} PRTE(F_\tau) = \int_0^1 MTE(u_s) \omega_{MP RTE}(u_s; \{F_\alpha\}) du_s \quad (13)$$

A practical issue in the estimation of the MP RTE concerns how to deal with the conditioning set of observed variables \mathbf{X} . Following Carneiro, Heckman and Vytlačil (2010), if the conditioning set contains only discrete elements the nonparametric estimation is still possible; however, as noted in Carneiro, Heckman and Vytlačil (2011) by imposing stronger assumptions, specifically invoking parametric assumptions on the joint distribution of the unobservables and independence between unobservables and (\mathbf{X}, \mathbf{Z}) , it is possible to identify the MTE over the unconditional support of P , not on the support of P conditional on \mathbf{X} .

An interesting objective that can be achieved through evaluation of public policy is performing a cost-benefit analysis. Carneiro, Heckman and Vytlačil (2010, 2011) state that the parameter of interest in this sense is the MP RTE. In particular, they link this parameter to the *Average Marginal Treatment Effect* (AMTE), defined as: “the average effect of treatment for the marginal person who is indifferent between participation and non participation”. Cost-

benefit analysis can be achieved comparing average marginal returns to average marginal costs of policy implementation.

AMTE relies on a certain metric; it considers individuals who are arbitrarily close to the margin of indifference, so it takes into account a measure of distance between $P(\mathbf{Z})$ and U_s that identifies the indifference set.

Given a certain metric $m(P, U_s)$, the AMTE is defined as follows:

$$AMTE = \lim_{e \rightarrow 0} E[Y_1 - Y_0 | m(P, U_s) \leq e] = \int_0^1 MTE(u_s) \omega_{AMTE}(u_s) du_s \quad (14)$$

Where weights, $\omega_{AMTE}(u_s)$, depends on the metric chosen.

Both MP RTE and LATE estimate marginal effects; LATE “measures the mean gross return to treatment for individuals induced into treatment by a change in an instrument”. It estimates “the mean return at the margin defined by manipulation of the instrument”. (see Heckman, 2010, 15). Anyway the substantial difference is that in the case of LATE the instrument variation has to correspond exactly to the policy variation and different instruments produces different estimates (e.g. exploiting variation over an instrument like distance to college does not produce the same estimate that would be obtained exploiting the variation in tuitions). Moreover, the LATE approach does not require the specification of a choice equation. This means that it is not possible to identify the margin of choice traced out by variation in instruments.

2.3 ESTIMATION’S PROCEDURE OF MTE AND MARGINAL POLICY EFFECTS

Heckman, Urzua and Vytlacil (2006) specify different estimation alternatives to derive the parameters of interest presented in the previous paragraph. In particular, the first step consist in the estimation of the propensity score; this is done by adopting a Probit model by authors, but other techniques can be adopted (e.g. the Logit model as in Carneiro, Heckman and Vytlacil, 2011).

The main point concerns the second step, that is the estimation of the MTE. Methods explained by Heckman, Urzua and Vytlacil (2006) can be pooled in two main categories: the parametric and the semiparametric approach of structural models; within each category, they specify two procedures, as listed below:

- *The parametric approach:*
 - Under normality assumption;
 - Relaxing normality assumption (using a polynomial for the propensity score);
- *The semiparametric approach:*
 - The LIV estimator;
 - Semiparametric approach with more structure (mixing polynomial approximation method and LIV).

Reconnecting to the model of Carneiro, Heckman and Vytlacil, (2011) previously discussed, they adopt a *Normal Selection Model* for what concerns the parametric estimation, and the LIV estimator afterwards, which is the main focus of their work. Under the first procedure, more structure is required with respect to the LIV estimator approach. In particular, the main parametric assumption invoked is that the joint distribution of (U_0, U_1, V) is normally distributed and independent of (\mathbf{X}, \mathbf{Z}) . Under this assumption, the outcome equation is estimated through Maximum Likelihood estimator. Authors argue that the parametric estimation is less flexible but more precise than the semiparametric one. Nevertheless, since normality is a strong assumptions, the results produced under this method are compared with that of the semiparametric approach.

In what follows, I discuss first the parametric approach and then the semiparametric one.

2.3.1 Sample Selection Model

The *Sample Selection Model* of Heckman was developed in Heckman (1979). In this section I discuss this topic, providing details and development of this method.

Heckman developed the model in order to estimate behavioural functions in case of “omitted variable” bias. This bias results from non-random selection of sample, therefore is also called *sample selection bias*. When non-random selection arises from individuals’ self-selection into treatment we talk about *self-selection bias*. Heckman (1979) developed this model to deal with cases of limited dependent variable. To explain this, consider for example a wage equation; when estimating it, it is possible to observe earnings only for a certain subgroup of the population, which is not randomly selected, but has self-selected on the base of some unobserved factors.

This model is particularly important in the heterogeneous treatment effect literature, because it has put the basis for the development of the estimation procedures needed for the identification of the MTE. What is particularly important in the estimation of the MTE using parametric or semiparametric approach instead of IV approach, is that we are able to precisely specify the margin of indifference.

Heckman's Nobel Lecture contributed to the further developments in the identification of the parameters of interests considered in policy evaluation. Heckman (1979) presented his techniques modelling labour supply to identify the determinants of wages of working women. For what concerns returns to schooling, through a structural model as the Generalized Roy model, it is possible to model both the outcome and choice equations as it has been explained in section 3.2.1. In particular, modelling schooling choice is important given endogeneity of the dummy variable indicating schooling status; here the selection bias arises because some determinants of the schooling choice equation affects the wage equation.

Heckman sample selection model relies on the use of the *Inverse Mill's Ratio* (see Heckman, 1979 for more details). It is implemented in the famous *Heckman two-step estimator* to estimate the outcome regression.

The sample selection model involves two equations, the outcome and the selection equation. In case of returns to education, the selection equation indicates that we are observing wages of graduates only if the selection equation is positive. This equation tells us how the selection process works.

Indeed, in the original sample selection model of Heckman, the selection equation defines that you can observe the dependent variable in the outcome equation only if the selection equation is above a certain threshold, e.g. is positive. For those individuals not selected into program/treatment/status you cannot observe the counterfactual. Therefore, the dependent variable in the outcome equation is *incidental truncated*, that means that it is the result of a sample selection mechanism⁷.

Now I present the main features of the sample selection model; this allow to understand the estimation procedures that are useful to our analysis.

Following Greene (2005) and Vella (1998), the sample selection model is structured as follows:

⁷ For more details about truncation and censoring see Greene (2005) and Heckman (1979).

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i; i = 1, \dots, N \quad (\text{i})$$

$$d_i^* = \mathbf{z}_i' \boldsymbol{\gamma} + v_i; i = 1, \dots, N \quad (\text{ii})$$

$$d_i = 1(d_i^* > 0) \quad (\text{iii})$$

$$y_i = y_i^* * d_i \quad (\text{iv})$$

y_i^* is the latent endogenous variable, whose counterpart is y_i which can be observed only if $d_i = 1$; d_i^* is the latent variable indicating the sample selection. \mathbf{x}_i' and \mathbf{z}_i' are vectors of exogenous variables and the exclusion restriction is that \mathbf{z}_i' contains at least one element not contained in \mathbf{x}_i' ; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of unknown parameters. ε_i and v_i are zero mean error terms and have nonzero correlation. The entire sample is made of N individuals and n is the number of individuals for which $d_i = 1$.

The standard assumption [A1] tells that, the error terms (ε_i, v_i) are bivariate normally distributed with correlation $\rho[(\varepsilon_i, v_i) \sim N(0, 0, 1, \sigma_\varepsilon, \rho)]$ and they are independent of \mathbf{z}_i' .

Under this assumption it follows that:

$$\begin{aligned} E[y_i^* | \mathbf{x}_i', d_i^* > 0] &= E[y_i^* | \mathbf{x}_i', d_i = 1] \\ &= \mathbf{x}_i' \boldsymbol{\beta} + E[\varepsilon_i | d_i = 1] \\ &= \mathbf{x}_i' \boldsymbol{\beta} + E[\varepsilon_i | v_i > -\mathbf{z}_i' \boldsymbol{\gamma}] \end{aligned} \quad (\text{v})$$

Since (ε_i, v_i) are bivariate normally distributed, we can know the truncated joint density given the truncated normal distribution formula from Greene (2003, 757). Given a continuous random variable x with pdf $f(x)$ and be a a constant, the density of the truncated random variable is:

$$f(x|x > a) = \frac{f(x)}{\text{Prob}(x > a)}$$

If x is normally distributed with mean μ and variance σ , $\text{Prob}(x > a) = 1 - \Phi(\frac{a-\mu}{\sigma})$, with $\Phi(\cdot)$ the standard normal cumulative distribution function. Therefore the density of the truncated normal distribution is:

$$f(x|x > a) = \frac{f(x)}{1 - \Phi(\frac{a-\mu}{\sigma})}$$

Given a random variable with a truncated normal distribution, the truncated mean is:

$$E(x|truncation) = \mu + \sigma\lambda(\alpha)$$

With $\alpha = \frac{a-\mu}{\sigma}$; $\lambda(\alpha)$ is the so called *Inverse Mills Ratio*, equal to:

- $\phi(\alpha)/[1 - \Phi(\alpha)]$ if truncation is $x > a$;
- $-\phi(\alpha)/\Phi(\alpha)$ if truncation is $x < a$.

Where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the probability density and the cumulative distribution function of the standard normal distribution.

Considering these formulas, Greene (2003), defines the moments of the *Incidentally Truncated Bivariate Normal Distribution*; if y and z have a bivariate normal distribution, with mean μ_y, μ_z , standard deviation σ_y, σ_z and correlation ρ , the truncated mean is:

$$E(y|z > a) = \mu_y + \rho\sigma_y \lambda(\alpha_z)$$

With $\alpha_z = \frac{a-\mu_z}{\sigma_z}$, and $\lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)]$.

Therefore, given the properties of the incidentally truncated distribution and in particular from the truncated mean formula, it follows that equation (v) can be written as:

$$\begin{aligned} E[y_i^* | \mathbf{x}_i', d_i = 1] &= \mathbf{x}_i' \boldsymbol{\beta} + \rho\sigma_\varepsilon \phi(-\mathbf{z}_i' \boldsymbol{\gamma}) / [1 - \Phi(-\mathbf{z}_i' \boldsymbol{\gamma})] \\ &= \mathbf{x}_i' \boldsymbol{\beta} + \rho\sigma_\varepsilon \lambda(-\mathbf{z}_i' \boldsymbol{\gamma}) \end{aligned} \quad (\text{vi})$$

From (vi) is clear that the least square regression of y_i on \mathbf{x}_i' in the observed data cannot allow to consistently estimate $\boldsymbol{\beta}$, this because of the second term in the right hand side of (vi) which is different from zero ($\rho\sigma_\varepsilon \neq 0$ for assumption) and because the inverse Mills ratio $\lambda(-\mathbf{z}_i' \boldsymbol{\gamma})$ is correlated with \mathbf{x}_i' in particular if the two vectors \mathbf{x}_i' and \mathbf{z}_i' contain common variables. Only introducing a term that accounts for $E[\varepsilon_i | d_i = 1] \neq 0$, it is possible to obtain a consistent estimate, otherwise one would incur in the *specification error* of an *omitted variable* problem.

An interesting extension of this model consists in the *Treatment Effect Model*, that is particularly interesting for the analysis presented in this dissertation and represents the setting shown in the paper by Carneiro, Heckman and Vytlačil (2011).

2.3.2 Endogenous Treatment Effect Model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \delta d_i + \varepsilon_i; i = 1, \dots, N \quad (\text{vii})$$

$$d_i^* = \mathbf{z}_i' \boldsymbol{\gamma} + v_i; i = 1, \dots, N \quad (\text{viii})$$

$$d_i = 1(d_i^* > 0) \quad (\text{ix})$$

This model indeed has been widely applied in the literature to study returns to education. Therefore, as shown in Greene (2003) consider equation (vii) as the earning equation and equation (viii) the program participation equation indicating whether the individual attends college or not. Since the framework is that of selection on gains, estimating equation (vii) with conventional methods like OLS would produce a biased estimate; indeed individuals who self-select into college are the ones with higher returns. Since $\boldsymbol{\beta}$ differs across individuals and it is in general found to be different from the average $\boldsymbol{\beta}$ for the population, $\bar{\boldsymbol{\beta}}$, then the *sorting gain* is defined as $E[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}} | S = 1]$.

When the sorting gain exists, the relation between the most known parameters estimated is $TT > ATE > TUT$, because those who decide to select into schooling are those who retrieve an higher returns from it.

Since v_i and ε_i are correlated, keeping the same assumption [**A1**] of the sample selection model, and keeping in mind equation (vi), the estimation turns to be:

$$\begin{aligned} E[y_i | \mathbf{x}_i', d_i = 1] &= \mathbf{x}_i' \boldsymbol{\beta} + \delta + \rho \sigma_\varepsilon E[\varepsilon_i | \mathbf{x}_i', d_i = 1] \\ &= \mathbf{x}_i' \boldsymbol{\beta} + \delta + \rho \sigma_\varepsilon \lambda(-\mathbf{z}_i' \boldsymbol{\gamma}) \end{aligned} \quad (\text{x})$$

Then, by symmetry of the normal distribution, $E[\varepsilon_i | \mathbf{x}_i', d_i = 0] = E[\varepsilon_i | v_i < -\mathbf{z}_i' \boldsymbol{\gamma}] = E[\varepsilon_i | v_i > \mathbf{z}_i' \boldsymbol{\gamma}] = \phi(\mathbf{z}_i' \boldsymbol{\gamma}) / [1 - \Phi(\mathbf{z}_i' \boldsymbol{\gamma})]$; so, for individuals that do not select into treatment,

$$\begin{aligned} E[y_i | \mathbf{x}_i', d_i = 0] &= \mathbf{x}_i' \boldsymbol{\beta} - \rho \sigma_\varepsilon E[\varepsilon_i | \mathbf{x}_i', d_i = 0] \\ &= \mathbf{x}_i' \boldsymbol{\beta} - \rho \sigma_\varepsilon \lambda(\mathbf{z}_i' \boldsymbol{\gamma}) \end{aligned} \quad (\text{xi})$$

Consequently, the average treatment effect is:

$$E[y_i | \mathbf{x}_i', d_i = 1] - E[y_i | \mathbf{x}_i', d_i = 0] = \delta + \rho \sigma_\varepsilon \phi_i / \Phi_i [1 - \Phi_i] \quad (\text{xii})$$

From (xii) it is clear that by simply estimating (vii) by OLS omitting $\rho\sigma_\varepsilon \phi_i / \Phi_i[1 - \Phi_i]$ produce a biased estimate of the true treatment effect.

2.3.3 Parametric estimation

The normal sample selection model and its extension as treatment effect model, can traditionally be estimated in two ways:

- Through the two-step estimator of Heckman (1979);
- Through Maximum Likelihood.

2.3.3.1 The two step estimator

The two- step estimator for the sample selection model involves:

- A first step in which a binary Probit model is applied to estimate the selection equation through maximum likelihood; this allow to estimate $\boldsymbol{\gamma}$. Since \mathbf{z}'_i is observed, using $\hat{\boldsymbol{\gamma}}$, it is possible to obtain an estimation of the inverse Mills ratio λ_i , therefore $\hat{\lambda}_i$.
- The second step consists in the least square regression of y_i on \mathbf{x}'_i and the additional term $\hat{\lambda}_i$. this regression takes the form of:

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \mu \hat{\lambda}_i + \eta_i \quad (\text{xiii})$$

With $\mu = \rho\sigma_\varepsilon$, and η_i a generic zero mean error uncorrelated with the regressor.

This method is also called *control function estimator*.

Given the formula of the variance of the incidentally truncated bivariate normal distribution (see Greene, 2003, 781):

$$\sigma_i^2 = \sigma_\varepsilon^2(1 - \rho^2 \delta_i), \quad \text{with } \delta_i = \lambda_i(\lambda_i - \mathbf{z}'_i \boldsymbol{\gamma})$$

It is possible to estimate all the parameters of the model.⁸

Finally, a t-test for $\mu = 0$ is a test for the sample selectivity bias.

In case of the endogenous treatment effect model, equation (xiii) becomes:

⁸Greene (2003, 2005) demonstrated that the covariance matrix for least square estimator as it is conventionally estimated is inappropriate in this model, anyway, Heckman (1979) provided the appropriately correction.

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \theta d_i + \mu \hat{\lambda}_i + \eta_i \quad (\text{xiii})$$

And the first step leads to obtain a *Probit residual* (estimation of λ_i using $\hat{\mathbf{y}}$), that is called the *generalized residual from the Probit model* and takes the form of:

$$\hat{\lambda}_i = d_i * \frac{\phi(\mathbf{z}_i' \hat{\mathbf{y}})}{\Phi(\mathbf{z}_i' \hat{\mathbf{y}})} + (1 - d_i) \frac{-\phi(-\mathbf{z}_i' \hat{\mathbf{y}})}{\Phi(-\mathbf{z}_i' \hat{\mathbf{y}})} \quad (\text{xiv})$$

From Vella (1998, 136), important properties of this residual are stated:

“First, it has mean zero over the whole sample. Second, it is uncorrelated with the variables that appears as explanatory variables in the first step Probit model. (...) This model is identified without exclusion restrictions due to the nonlinearity of the residual. Also note that the generalized residual is uncorrelated with the \mathbf{z}_i' s, over the whole sample, by construction. Thus the consequences of high degree of collinearity between the generalized residual and the \mathbf{z}_i' s, which is a concern in the sample selection model, does not arise”.

The two-step estimator has also been revised in the literature in order to accommodate criticism about the strong distributional assumption imposed. Therefore, as explained in Vella (1998) and Greene (2003), alternative methods like the *Semi-parametric two-step estimation* has been investigated. See also Heckman, Urzua and Vytlačil (2006) that suggest the adoption of a parametric approach using a polynomial approximation for the propensity score to estimate the marginal treatment effect under heterogeneity conditions.

2.3.3.2 The Maximum Likelihood estimator

Under assumption [A1] the average log likelihood function to maximize is:

$$L = \frac{1}{N} \sum_{i=1}^N \left\{ d_i * \ln \left[\int_{-\mathbf{z}_i' \boldsymbol{\gamma}}^{\infty} \phi_{\varepsilon v}(y_i - \mathbf{x}_i' \boldsymbol{\beta}, v) dv \right] + (1 - d_i) * \left[\ln \int_{-\mathbf{z}_i' \boldsymbol{\gamma}}^{\infty} \int_{-\infty}^{\infty} \phi_{\varepsilon v}(\varepsilon, v) d\varepsilon dv \right] \right\} \quad (\text{xv})$$

With $\phi_{\varepsilon v}$ the probability density function of the bivariate normal distribution.

From Vella (1998), “when the model is estimated by maximum likelihood the parameter estimates are fully efficient”. Also in this case, there have been various attempts in the literature to estimate under ML relaxing normality assumption; see the discussion in Vella (1998) for more details.

Carneiro, Heckman and Vytacil (2011) estimate the normal selection model through Maximum Likelihood. Under the assumption that the joint distribution of (U_0, U_1, V) is normally distributed and independent of (\mathbf{X}, \mathbf{Z}) with the variance of V normalized to 1, they adopt a linear-in-the-parameter model, assuming separability between \mathbf{X} and (U_0, U_1) , with equations (2) and (7) becoming:

$$Y_1 = \delta_1 \mathbf{X} + U_1 \text{ and } Y_0 = \delta_0 \mathbf{X} + U_0 \quad (2')$$

$$I_s = \gamma \mathbf{Z} - V \quad (7')$$

Therefore, writing equation (2') with potential outcome notation, we obtain:

$$Y = \delta_0 \mathbf{x} + S \mathbf{x} (\delta_1 - \delta_0) + U_0 + S(U_1 - U_0) \quad (5')$$

Carneiro, Heckman and Vytacil (2003), show that combining the model for S with the model for Y implies a partially linear model for the conditional expectation of Y :

$$E(Y|\mathbf{X} = \mathbf{x}, P(\mathbf{Z})) = \delta_0 \mathbf{x} + P(\mathbf{Z}) \mathbf{x} (\delta_1 - \delta_0) + K(P(\mathbf{Z})) \quad (8')$$

Where $K(P(\mathbf{Z})) = E(U_1 - U_0 | P(\mathbf{Z}), S = 1) P(\mathbf{Z}) = E(U_1 - U_0 | \Phi(U_s) \leq P(\mathbf{Z})) P(\mathbf{Z})$, and $(\delta_1 - \delta_0)$ is the coefficient of the interaction between $P(\mathbf{Z})$ and \mathbf{x} . In Carneiro, Heckman and Vytacil (2003) testing the linearity of this equation is a way to test for selection on the individual returns to attending college; Nonlinearity in P means that there is heterogeneity in the returns to college attendance and selection on gains (conditional on \mathbf{X}).

Deriving (8') with respect to p , we are able to estimate the marginal treatment effect.

$$\begin{aligned} \text{MTE}(\mathbf{x}, u_s) &= \mathbf{x}(\delta_1 - \delta_0) + E(U_1 - U_0 | U_s = u_s) \\ &= \mathbf{x}(\delta_1 - \delta_0) + E(U_1 - U_0 | V = \Phi^{-1}(u_s)) \\ &= \mathbf{x}(\delta_1 - \delta_0) - (\sigma_{1V} - \sigma_{0V}) \Phi^{-1}(u_s) \end{aligned} \quad (10')$$

With, $\Phi^{-1}(\cdot)$ the inverse of the standard normal cumulative distribution function. A test for heterogeneous effects require to test if the slope of the MTE is equal to zero, which means, testing if $(\sigma_{1V} - \sigma_{0V}) = 0$.

2.3.4 The LIV estimator

The innovation of Heckman and Vytlacil (1999) allows to exploit a semiparametric estimation of the parameters of interest that overcomes the concerns raised up about the strong parametric assumptions required in the previously seen approach. As explained in Carneiro, Heckman and Vytlacil (2001, 3) “The contrast often made in the empirical literature between IV and selection models is a false one. Recently developed IV methods are special cases of nonparametric selection models”.

Start by estimating $P(\mathbf{Z})$ through Probit or Logit regression, as we have seen before. This step allows to identify the support of $P(\mathbf{Z})$ on which MTE will be estimated. Depending on the underlying assumptions, it will vary. Indeed, assuming that (U_0, U_1, V) is independent of \mathbf{Z} given \mathbf{X} , Carneiro, Heckman and Vytlacil (2011) show that the support of the propensity score shrinks with respect to the full unit interval⁹. Whereas, invoking the stronger assumption for which (U_0, U_1, V) is independent of (\mathbf{Z}, \mathbf{X}) , the support is almost the full unit interval.

The subsequent step consist in using $\widehat{P(\mathbf{Z})}$ to estimate a partially linear regression of Y on \mathbf{X} and $P(\mathbf{Z})$. This step allow to identify δ_1 and δ_0 .

Estimating a partially linear regression under the assumptions of separability and independence between \mathbf{X} and unobservables, presents the important advantage of relying only on the marginal support of $P(\mathbf{Z})$, instead of investigating the support of $P(\mathbf{Z})$ conditional on \mathbf{X} .

The term $K(P(\mathbf{Z}))$ in equation (8') is an unknown function that must be estimated non parametrically. Carneiro, Heckman and Vytlacil (2006, 2011) use a local polynomial estimation for the estimation of $K(P(\mathbf{Z}))$ and its derivative with respect to $P(\mathbf{Z})$, necessary for the identification of the MTE. Their approach suggests: from equation (8'),

$$K(P(\mathbf{Z})) = E(Y - \delta_0 \mathbf{x} + P(\mathbf{Z})\mathbf{X}(\delta_1 - \delta_0) | P(\mathbf{Z}))$$

From which, $K(P(\mathbf{Z}))$ results from the local polynomial regression of $Y - \widehat{\delta_0} \mathbf{x} + \widehat{P(\mathbf{Z})}\mathbf{X}(\widehat{\delta_1} - \widehat{\delta_0})$ on $\widehat{P(\mathbf{Z})}$. Carneiro, Heckman and Vytlacil (2003) instead estimate $K(P(\mathbf{Z}))$

⁹ Given that \mathbf{X} is multidimensional they consider an index as: $\mathbf{X}[\delta_1 - \delta_0]$.

through local linear regression. Carneiro (2003) estimate equation (8') using both approaches, a local linear regression using a biweight kernel and then polynomials in P .

Details about semiparametric LIV estimator estimation are presented in Heckman, Urzua and Vytlačil (2006).

3 FUNDING HIGHER EDUCATION IN THE U.K.

Investments in human capital is a major concern for governments of all countries. Designing reforms aimed at promoting education at all levels is an activity that could focus on broadening the participation to education among citizens and/or enhance the efficiency of the services provided, for instance by ameliorating the quality of teaching.

In what follows, I introduce the functioning of the higher education system in the U.K., focusing on the aspect of the costs of access and the related reforms.

3.1 HIGHER EDUCATION REFORMS IN THE U.K.: FROM 1960S TO OUR DAYS.

Data from the *Higher Education Statistics Agency* (HESA) show Higher Education (HE) students' enrolments from the academic year (a.y.) 2009/10 to 2013/14; considering undergraduates students, enrolments remained more or less stable until a.y. 2011/12 followed by a large decrease of 6 percentage points in 2012/13, in correspondence to changes in tuition fees. (The fee's cap has been increased to £9,000 for new entrants from 2012/13)¹⁰. The decline continued by another 2% between 2012/13 and 2013/14.

At first glance, an overview of the situation over the last years catches the attention on the relationship between university enrolment and tuition fees/grants changes over time.

Wyness (2010) states that in the last 50 years, students' volume in higher education sector in the U.K. has more than quadrupled. Nonetheless, "in the late 1980s, the U.K. had one of the lowest participation rates in higher education (about 14 per cent) of any advanced industrial country". (Barr and Crawford, 2005).

The first stylized fact of this phenomenon is that while students' volume were rising, funding to higher education sector were decreasing. The second stylized fact concerns social characteristics of students. There are much more students coming from wealthy families as compared to students from more disadvantaged socio-economic conditions. (Wyness, 2010).

Increasing funding and stimulating participation, in particular among less wealthy people, is a goal that can be achieved through targeted reforms.

¹⁰ Source: <https://www.gov.uk>

Therefore, I illustrate below the major interventions realized by the U.K.'s Government to address these issues.

Today, funding system in the U.K. provides that HEIs (*Higher Education Institutions*) are publicly funded through:

- tuition fees backed by Government-funded loans;
- teaching grants from the *Higher Education Council for England*.

For students coming from least wealthy backgrounds, the Governments also provide means-tested maintenance loans and maintenance grants¹¹. This is the results of several reforms succeeded over the last 50 years.

The major policy changes happened over 5 decades are described below. The last relevant HE reforms happened in 2012.

3.1.1 1960s – 1990s

Until 1963 in the U.K., policies in force envisaged a system in which all costs related to university, from teaching to tuition fees, from grants to administrative costs, were borne by taxpayers. Funding per student was high but the volume of students enrolled was small. For this reason, in order to encourage participation, in 1963 the *Robbins Report* led to an expansion in the number of universities. This expansion in the following twenty years pushed the volume of students up, even if it was still low as compared to the other industrial countries. Moreover, being part of the higher education system still was a privilege of richer families. The funding system showed some weaknesses given that maintenance grants (non-repayable financial support) were entirely paid by the Government; for this reason, student loans entered the funding system.

From 1990s to our days several reforms modified Higher Education's access costs. In the following tables (*Table 1- 3*) relevant values (nominal) of grants, fees and loans are reported with respect to subsequent income brackets for academic years in which changes took place.

¹¹ See *Government reform of higher education: twelfth report of Session 2010-12*, (2012).

Table 1. Grants by parental income (nominal value)

<i>GRANTS</i>	Academic Year				
Parental income	1992/93	1998/99	2004/05⁴	2006/07⁵	2012/13⁶
≤10,000	2265 ¹	810 ³	1000	2.700	3.250
20000	136 ²	810 ³	298	2.283	3.250
30000	0	456	0	832	2.341 ⁷
40000	0	0	0	0	523
50000	0	0	0	0	0
≥60000	0	0	0	0	0

Source: computation made by Erich Battistin

[1] Statistical First Release 1992-2002

[2] Computed from Dearden et al (2011)

[3] The Independent (1998), Student Choice

[4] A Guide to Financial Support for Higher Education Students in 2005/06

[5] Dearden et al (2011)

[6] Statistical First Release 2012/13

[7] $3250 - \frac{30000-25000}{5.5}$, Reduced by £1 for every £5.50 of income above £25,000 up to £42,600

Table 2. Fees by parental income (nominal value)

<i>FEES</i>	Academic Year				
Parental income	1992/93	1998/99	2004/05²	2006/07	2012/13
≤10,000	0	0	0	3,000	9,000
20000	0	280 ¹	0	3,000	9,000
30000	0	1,000	886 ³	3,000	9,000
40000	0	1,000	1,150	3,000	9,000
50000	0	1,000	1,150	3,000	9,000
≥60000	0	1,000	1150	3,000	9,000

Source: computation made by Erich Battistin

[1] The Independent (1998), Student Choice

[2] A Guide to Financial Support for Higher Education Students in 2005/06

[3] $\frac{30000-22010}{9.5} + 45$

Table 3. Loans by parental income (nominal value) including Fee Loans introduced in 2006/07

LOANS Parental income	Academic Year				
	1992/93 ¹	1998/99	2004/05 ⁴	2006/07	2012/13 ⁶
≤10,000	580	2,735 ²	4,095	6,555	12,875
20000	580	2,735 ²	4,095	6,555	12,875
30000	580	2,326 ³	4,095	7,005	13,330 ⁷
40000	580	2,051 ²	3,331 ⁵	6,459	14,239
50000	580	2,051 ²	3,070	6,305	13,770
≥60000	580	2,051 ²	3,070	6,305	12,575 ⁸

Source: computation made by Erich Battistin

[1] Statistical First Release 1992-2002

[2] Student Loans - A Guide to Applying in 1998/99

[3] Computed from Dearden et al (2011)

[4] A Guide to Financial Support for Higher Education Students in 2005/06

[5] $3075 + 1020 - \left(\frac{40000 - 22010}{9.5} + 45 - 1175 \right)$. This formula is derived from *A Guide to Financial Support for Higher Education Students in 2005/06*, but the figure for the maximum grant is from *Statistical First Release 2004/05*.

[6] Statistical First Release 2012/13

[7] "The amount of Maintenance Grant you receive will affect the amount of Maintenance Loan you can borrow. We will reduce the amount of Maintenance Loan you can receive by £0.50 for every £1 of Maintenance Grant you are entitled to". $(3250 - 2341)0.5 + 3875 + 9000$

[8] All students are entitled to 65% of the appropriate maximum Maintenance Loan, but the remaining 35% is subject to means-testing.

3.1.2 1990s – 2000s

In 1990 the *Student Loans Company* was funded and the first *Student Loan System* was implemented. Nonetheless, the way in which loans' accounting rules were set created some funding problems. In 1992 the *Further and Higher Education Act*¹² was issued by the Government in order to convert a certain number of polytechnics and colleges of higher and further education into universities, and create bodies to fund higher education. In 1997 in order to manage the funding crisis, the *Dearing Report* was issued. Indeed, the expansion in the number of universities and students volumes was not sustained by a correspondent increase in the available funding per student; for this reason, the *National Committee of Inquiry into Higher Education* (chaired by Lord Dearing) was established in order to manage the situation. The *Dearing Report* introduced an important change because it involved that new full-time students enrolling in the a.y. 1998/99 started to contribute to the costs of higher education. The 1998 reform was therefore a turning point. Indeed, in 1998 the *first tuition fees*

¹² See <http://www.legislation.gov.uk/ukpga/1992/13/contents>

were introduced officially for the a.y. 1998/99 through the *Teaching and Higher Education Act*; the fees amounted to £1,000. This was paid up-front just by richer families while poorest were exempted, given that the amount to be paid was contingent on student's and parents' income.

Maintenance grants were abolished in 1999. For what concerns maintenance loans, these were increased by an amount similar to that by which grants decreased and fees increased. The objective of the government was to try to leave unaltered economic conditions of students in the period post-reform.

3.1.3 2001 – 2015

In 2004 the *Higher Education Act* was issued. It abolished up-front fees, introduced a deferred fee to be implemented in 2006/07 a.y. (in England and Northern Ireland and in 2007/08 in Wales), and re-introduced grants for a.y. 2004/05 at £ 1,000 per year for low income families.

The main change produce by the *Higher Education Act* consisted in the introduction in the a.y. 2006/07 of the “deferred variable fee” (or “top-up fees”) not means-tested; each university could decide the amount respecting the cap of £3,000. Under this reform then, students face a “tuition-free entry” and start paying fees just after graduation. In order to promote participation, universities had to develop the *Access Agreement* with the *Office for Fair Access* (OFFA) in order to establish measures to support students' participation, like bursary and other measures.

Fees are with this reform deferrable until after graduation through government-subsidies *Tuition Fee Loans*, issued at zero real interest rate and repayable according to income. The characteristics of *Tuition Fee Loans* (in term of interest rate and repayment term) are the same of *Maintenance Loan* and both can be combined.

As for the previous reform in 1998, indeed, increase in fees has been balanced by the loans and grants. This reform did not exempt poorest students from fees' payment but universities charging fees of more than £2,700 had to offer bursaries of at least £300 to students receiving the maximum amount of *Maintenance Loan*¹³.

¹³ See Tuition Fee Statistics (2015)

The *Browne Report*¹⁴ (*Securing a Sustainable Future for Higher Education: An independent review of higher education funding and student finance*) published in October 2010, was aimed at achieving three objectives:

- increase the overall higher education participation, in particular, extending participation to poorer students that could not afford higher education;
- improve teaching quality to enhance students' knowledge needed in the labour market;
- simplify the funding system.

In June 2011 the Government's *Higher Education White Paper: Students at the heart of the system*¹⁵ was published. It rejected some *Browne Report*'s proposals while accepting some others. The results are indicated below and entered into effects in the a.y. 2012/13.

Following the *Browne Report*, the Government increased the cap on tuition fees from a 'basic maximum amount' of £6,000, to an absolute maximum of £9,000 which could be charged only in 'exceptional circumstances'. Students would be entitled to tuition fee loans of up to £9,000 per year, according to the fees charged by the institution they attend. These changes started to be applied from the a.y. 2012/13 for undergraduates. This is the current situation.

Tuition fee Loans are still available to cover fees for both full and part-time students. Data from the *House of Commons* (see Bolton, 2015) tells that for the a.y. 2012/13 universities set their fees at £8,400 yet. For the current a.y. 2015/16 they increased fees up to almost £8,900. In the same year, about 92% of eligible full-time students from England took out *Tuition Fee Loans*.

In order to face living expenses, full-time students have also access to maintenance loan, which are 65% not means-tested (students will be entitled for at least 65% of the maximum loan¹⁶) and the remaining 35% is household's income contingent.

Moreover, grants were established up to £3,250 and bursaries and fee waivers were also available.

¹⁴ See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/422565/bis-10-1208-securing-sustainable-higher-education-browne-report.pdf

¹⁵ See <https://www.gov.uk/government/publications/higher-education-students-at-the-heart-of-the-system--2>

¹⁶ See <http://www.publications.parliament.uk/pa/cm201213/cmselect/cmbis/286/286.pdf>

Scottish HE policies have diverged under some aspects with respect to the rest of the U.K.; from the a.y. 2000/01 upfront tuition fees were abolished for eligible *full*-time students. From 2001/02 the *Graduate Endowment* came into effects, that consisted in a contribution (of £2,000 in that year) made after graduation to be repaid in the same way as income-contingent loans. In 2007 the *Graduate Endowment* was abolished.

3.2 COLLEGE PARTICIPATION: THE EFFECTS OF THE REFORMS AND OTHER INFLUENCING FACTORS

Dearden, Fitzsimons, Wyness (2014) use a difference in difference approach to test the impact of the re-introduction of grants in the a.y. 2004/05 for students coming from poorer families, discovering that a £1,000 increase in grants produces an increase in participation among less well-off individuals of almost 4 percentage points.

The motivation of this analysis arises from substantial differences in applications rates among young people with different social backgrounds observed over the years. Indeed, data from UCAS (2012) confirm this evidence; application rates vary a lot with respect to backgrounds defining a gap between young people coming from different areas. Even if application rates of students coming from most disadvantaged areas have sharply increased (about 60%) between 2004 and 2012 contributing to the reduction of the gap, a large differential is still present: “Those living in the lowest income areas have application rates in 2012 of 23 per cent compared to 48 per cent for their peers living in the highest income areas”. (UCAS Analysis and Research, 2012).

Dearden, Fitzsimons and Wyness (2011) estimate the effects of grants and fees on individuals’ likelihood of entering university considering eligible students over the period 1992-2007. They created a pseudo-panel dataset estimated through a fixed-effects model with cohort defined aggregating observations on the base of geographic residence, gender, parental education and time. Results show that an increase of £1,000 in fees determines a decrease of 3.9% point in the likelihood of enrolling into university. An increase in grants of the same amount increase participation of 2.6 percentage point.

Two important aspects are underlined in this analysis: the impact of parents’ education and prior individual’s educational attainments are strong and significant. In particular, individual

having more educated parents and with good secondary education results are more likely to enrol into university.

Social background, family's financial situation, cognitive and non-cognitive skills are all important aspects to be considered when performing empirical analysis in the educational sector and, above all, should be factors to be investigated in order to develop effective policy reforms.

What emerges from data and research results discussed above is that financial aid to more disadvantaged individuals has a positive causal effect in terms of higher education participation but there also other factors driving educational choices that need to be addressed. Therefore, we can ask ourselves: are tuition alone sufficient to reduce the existing gap and lead to a future convergence between different socio-economics groups of individuals? Or are there other ways to efficiently invest Government resources?

When considering that in order to be accepted into university a student has to obtain some further educational qualification¹⁷ (e.g. *A levels* or other qualifications depending on the university chosen) it could be argued that reforming university's access costs would not be sufficient for these individuals when they have not even achieved the minimum requirements to enter it.

Individuals may find themselves in this situation as a consequence of the socio-economic background of their family. Obviously, other factors determine education achievement, in particular cognitive and non-cognitive skills but this does not preclude that there exists a category of children with the suitable ability, whose performances and ambitions are restraint by the environment and family.

Crawford and Greaves (2015) have highlighted these aspects as the results of their research. They investigate the effect of socio-economic background on university. Raw differences between highest and lowest socio-economic quintile groups are large (the first have 3 times more probability of enrolling than the other); the interesting aspect is that differences are reduced a lot when controlling for prior education attainment. In particular, results obtained by children at 16 years old (*General Certificate of Secondary Education*,

¹⁷ Education system in the U.K. involve children to be in secondary education (compulsory) until 16 years old. Next stages are Further Education (FE) from 16 to 18 years old and Higher Education (HE) which involve undergraduate and postgraduate courses.

GCSE) are alone sufficient to explain a substantial part of raw differences. Secondary school seems to play a crucial role even because it influences post-secondary school achievement (Further Education). The latter, most of the times necessary to entry into university. The evidence suggests that children from more disadvantaged families get worse secondary school results.

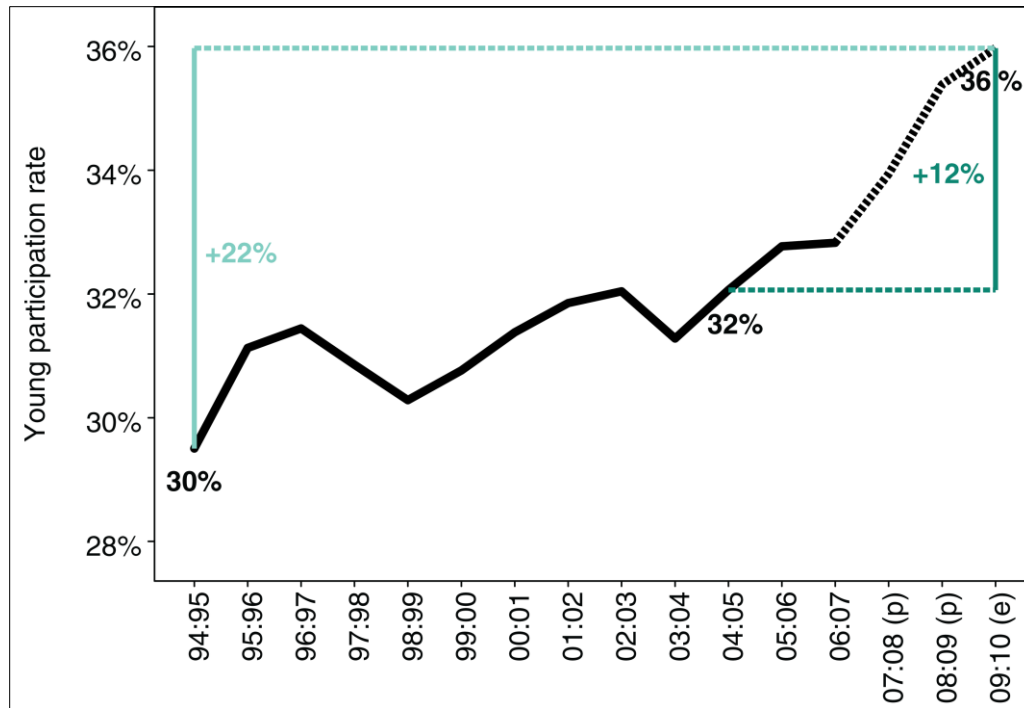
Chowdry et al., (2012) confirm this evidence: “(...) poor attainment in secondary schools is more important in explaining lower HE participation rates amongst students from disadvantaged backgrounds than barriers arising at the point of entry into HE”. This leads to investigate the true nature of constraints that prevents enrolment in higher education. Maybe there are several factors that induce individuals to stay out of higher education other than economics issues, that in any case are an important obstacles faced by more disadvantaged families. Policy reforms that address this concerns are necessary in inducing individuals into participation, nevertheless as Chowdry et al. (2012) argue, most disadvantaged students my anticipate access barriers to the time they are in secondary education precluding themselves the possibility of obtaining a degree, besides to whichever tuition reforms could be put in place.

Data from Corver (2010), reports trends in the enrolment rate into higher education institutions in England.

As shown in *Figure 1* the volume of young people enrolling into higher education sector has continued to increase starting from mid-1990s. In particular, we can see that after the reform for the a.y. 1998/99 that introduced for the first time tuition fees the pattern has not been negatively influenced. The report specifies that starting from the mid-2000s the differences in participation between young people from different backgrounds started to contract; the 2003/04 reform actually produced more favourable access condition for poorer students while establishing more burdensome rules for richest families¹⁸. Indeed, even if the proportion of students from more advantaged backgrounds has increased by 5 per cent over last five years, proportion of students from more disadvantaged ones has increased by 30 per cent over the same period in England.

¹⁸ Disadvantaged backgrounds are defined on the base of disadvantaged neighbourhoods, parental education, occupation or income.

Figure 1. Trends in young participation in England



Source: Trends in young participation in higher education: core results for England (2010)

The report, even if it does not provide evidence on the existence of any causal effect, underlines the influence that prior educational attainment has on the trend described above: “The increases in the proportion of young people living in the most disadvantaged neighbourhoods who enter higher education are consistent with other statistics including recent trends in GCSE attainment”. (See Corver, 2010, 2).

All these aspects considered together, suggests that there is still a lot to investigate in order to understand the mechanism that shapes individual decision-making process. The focus should be understand what are the main drivers of higher education, how to intervene on these ones to stimulate university participation and identify the target population that should be the focus of future reforms.

4 ESTIMATING RETURNS TO HIGHER EDUCATION IN THE U.K.

In this chapter I will present the analysis conducted on U.K. data to estimate the effect of holding an higher education qualification on future earnings.

The aim is that of evaluate the monetary returns that individuals induced to enrol into higher education by a set of policy reforms have experienced, and possibly perform an ex-ante evaluation that allows to considering potential efficient changes in future policies.

4.1 SAMPLE SELECTION

The analysis conducted in this dissertation is based on data from the *British Household Panel Survey* (BHPS). It consists in a panel dataset that follows a sample of individuals over time, from 1991 to 2009. It is made of 18 waves, each one reporting data at an individuals and household level. With annual frequency, each adult member (aged at least 16 years old) of a household of a nationally representative sample of more than 5,000 households is interviewed, resulting in approximately 10,000 individual interviews. When an individual left the initial household, it was followed to the new one interviewing the new components.

For the purpose of this dissertation this dataset is useful to detect all information at the individual level over time and also link them to information at the household level related to each individual selected.

The aim is evaluating the returns to higher education (focusing on undergraduate degrees) for a sample of individuals that face certain levels of university's access costs and benefits at the time they are eligible.

I have developed a cohort analysis considering eleven cohorts of individuals selected with the following procedures:

- For each year between 1992 and 2002 individuals “eligible” for university's enrolment has been considered. Eligibility is a condition based on the date of birth: “(...)in the UK, eligibility for the first year of HE is determined by date of birth. (...) youths become eligible for HE if they are aged 18 before August 31st of that academic year. This means that young people can be aged either 18 or 19 when they first become eligible for HE”. (See Dearden, Fitzsimons and Wyness, 70,

2014). Therefore, cohorts are created considering the date of birth such that for each cohort individuals are already aged either 18 or 19.

- The second step consists in looking for each individual at their wages 6 years after the enrolment (therefore 3 years after graduation).
- Finally, all other relevant explanatory variables are selected.

In what follows I list and described dependent and independent variables considered in the equations to be estimated.

4.2 DEPENDENT AND INDEPENDENT VARIABLES

Consider the *Mincer Equation* as presented in *Chapter 2*:

$$Y_i = \alpha + \beta S_i + \varepsilon_i \quad (1)$$

The dependent variable and the endogenous regressor are:

- *Logpaygu*: logarithm of *wpaygu* “usual gross pay per month of the current job”. This variable has been preferred to other earning’s indicator because of the limited number of missing values.
- *Schooling*: the treatment status dummy has been constructed considering the variable *wqfedhi* (highest educational qualification) in the BHPS. This variable has been observed for all individuals selected, in the year in which their wages are measured. This is the only indicator that permits to determine if an individual has achieved some qualifications from 18 years old onward. Other variables denoting educational achievements are also present, but they are uninformative because of the high rates of missing values.

Therefore, all individuals claiming to have a “First Degree”, “Teaching Qualification”, “Nursing Qualification” or “other Higher Qualifications¹⁹” (Higher Degree²⁰ excluded) are selected in the treatment group. 808 individuals are observed in total, 486 for S=1, 322 for S=0.

¹⁹ Foundation Degree, Diploma of Higher Education, Certificate of Higher Education, Higher National Diploma. (See <http://www.educationuk.org/global/articles/higher-education-courses-qualifications/>)

²⁰ Postgraduate qualifications

Hereafter, I estimate returns to Higher Education adopting methodologies explained in *Chapter 2.*, in order to deal with *selection in levels* and *selection on returns*. Consequently, I define a set of instrumental and control variables, as detailed below.

Instrumental Variables:

- Z is the vector of instrumental variables:
 - *Fees, Grants, Loans*: these are the access costs and benefits for a.y. between 1992/93 – 2002/03 established by the reforms presented in *Chapter 3*. They are attached to each individual on the base of the household income selected in the year the students is eligible. The variable considered to determine family income is *wfihhyr* (annual household income from September previous here to September current year).
 - *Distance*: the variable indicates the distance to the nearest university in miles. This variable was not available in the BHPS and it has been constructed. BHPS data of the Local Authority Districts (LAD²¹) codes have been selected for each household in which the individual was living at eligible age. For the limited number of waves it was not possible to identify the LAD of residence at 16 years old. From the higher education statistics agency website²², I have retrieve data about *Higher Education Providers*. Using geographic coordinates of LAD and Universities I have calculated distance in miles to the nearest institute. Distance to the nearest university is typically considered as an instrument because it should affect the costs faced by the individual that attend university.

Control variables:

- X is the set of controls that forms the conditioning set:
 - *Sex*
 - *Age*

²¹ LAD definition refers to the “Census 1991: Individual Sample for Anonymised Records for Great Britain (SARs)”.

²² <https://www.hesa.ac.uk/component/heicontacts/>

- *Qual*: this dummy indicate if the individual declares to possess GCSEs with grade A-C²³ (or equivalent qualification for Scotland: Standard Grades). A more suitable indicator of educational attainment would be the number of subjects passed with highest grade (e.g. at least 5 GCSEs grade A-C), unfortunately even if present, these variables report a high number of missing values. As discussed in *Chapter 3*, qualifications obtained at the end of secondary school influence the probability of attending university.
- *Jbstat*: is a categorical variable (recoded with respect to the original present in the BHPS) that indicates if in the eligibility year the individual is either working (1), studying (2) or find itself in other status (e.g. is unemployed, in family care, disabled) (3).

The decision of controlling for this variable derives from the fact that an individual who is already employed at 18 years old may be less motivated to enrol, if he finds itself in a good working environment or, alternatively, it could also be that having a job stimulates the propensity to enrol since the individual would feel itself more able to face its own spending needs, weighing less on parents.

- *Mqfedhi*: highest educational mother qualification reported in the year the individual is eligible. Since this variable was not present in the dataset it has been constructed keeping the identification number of the mother of each individual and looking at its level of educational qualification reported in the “eligibility” year of the children. For limited number of waves available, it was not possible to retrieve the educational qualification of the mother when the individual was in secondary school. It takes value 1 for “further/higher qualifications”, 2 for “secondary education or apprenticeship”, 3 for “no qualifications”. The importance of introducing this variable among the controls is deducted from the impact that familiar characteristics have on the decision of children to enrol, as explained in *Chapter 3*.

- *Fisit*: it is a subjective indication of the financial situation expressed by the father of the household. It has been re-managed from the original variable in the dataset, keeping the identification number of the father and looking at the variable

²³ *The General Certificate of Standard Education* is a qualification obtained at the end of compulsory education (15-16 years old) throughout all the country with standardized grades between A and G. Good qualifications, also required by colleges and universities, are those graded A-C.

fisit in the year of “eligibility” of the children. It is a dummy that takes value one for “wealthy/relaxed” situations and zero for “borderline/bad” situations. The reason for introducing it in the dataset is the same concerning the importance of controlling for familiar backgrounds.

- *Unat18*: is the unemployment rate for young aged 15-19 years old in the U.K. for all educational levels, measured in the year individual are eligible. The source is the *Eurostat LFS*²⁴. This variable is used in the conditioning set of the first stage regression as a measure to control for the characteristics of the labour market at the time the individual has to choose if enrol into Higher Education or look for a job.
- *Unwork*: is the unemployment rate for young up to 25 years in the U.K. measured in the year in which wages are observed. The source is the *Eurostat LFS*. This variable is used in the conditioning set of the outcome equation as a measure to control for labour market characteristics at the time the individual is already graduated.
- *Incomeclass*: this is a categorical variable that indicates income brackets over which fees, grants and loans are calculated. It takes value 1 for “low income” (<£20.000), 2 for “medium income” (£20.000-£40.000), 3 for “high income” (>£40.000). It is used in the conditioning set of the first stage regression.
- *Cohort dummies* and *regional dummies*. Regions considered are those indicated by the variable *Region2* (Government Office Region)²⁵.

Table 4. shows the descriptive statistics of the variables presented above. It is clear that certain variables determine a great reduction in the number of observations. This refers in particular to *mqfedhi* and *fisit*, and as explained later this can be a source of bias in the estimation.

²⁴ See <http://ec.europa.eu/eurostat/data/database>

²⁵ North East, North West, Yorkshire & Humber, East Midlands, West Midlands, East of England, London, South East, South West, Wales, Scotland.

Table 4. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
logpaygu	808	7.071698	0.5100139	4.685213	8.613273
schooling	808	0.6014851	0.4898957	0	1
fees	808	312.4072	459.3291	0	1075
grants	808	692.7327	938.2617	0	2265
loans	808	1971.601	1103.971	580	3905
distance	784	7.740855	7.73734	0.03	38.31
sex	808	0.480198	0.4999172	0	1
age	808	18.13738	0.3444573	18	19
qual	796	0.0942211	0.2923198	0	1
jbstat	807	1.656753	0.666386	1	3
mqfedhi	471	1.876858	0.7947404	1	3
fisit	444	0.6509009	0.4772231	0	1
unat18	808	16.40718	1.739745	13.5	19.2
unwork	808	12.9724	1.054673	11.7	15
incomeclass	808	1.851485	0.7301086	1	3

4.3 MODEL'S EQUATIONS

Marginal returns to a Higher Education are estimated here applying a parametric normal model using the Heckman two-step procedure.

Recalling the theoretical framework defined in [section 2.3](#) and Equations (2)-(7) describing the Generalized Roy Model, in *Chapter 2.*, the estimation involves the following steps²⁶:

1. Estimation of the Propensity Score from the first stage regression:

$$P(\mathbf{z}) = \Pr(S = 1 | \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x})$$

From this regression it possible to obtain the predicted values of μ_s , $\widehat{\mu}_s$, the coefficients from the regression of the schooling indicator on the instruments conditional on the set of controls. The predicted value of the propensity score is therefore $\widehat{P}(\mathbf{z}) = \Phi(\widehat{\mu}_s \mathbf{Z})$. Calculate then the normal density function using $\widehat{\mu}_s$, equal to $\phi(\widehat{\mu}_s \mathbf{Z})$, that allows to generate the selection term (Inverse Mill's ratio).

²⁶ See Heckman, Urzua and Vytlačil (2006) and Heckman, Tobias and Vytlačil (2001).

2. At this point is necessary to define the common support (between 0 and 1) of the propensity score. Select intervals of 0.01 points to determine the grid over which to compare the frequencies of the propensity scores for both groups of treated and untreated. For each range, keep observations only if positive frequencies in both group ($S=0,1$) exist. The marginal treatment effect as evaluated here below makes sense only within this common support.

3. Run the outcome regressions for each group separately, thus:

$$E(Y|X = x, S = 1, P(Z) = p) = \widehat{\alpha_1 + \varphi} + X\beta_1 + \rho_1 \left(-\frac{\Phi(\widehat{\mu_s Z})}{\Phi(\widehat{\mu_s Z})} \right) \text{ for } S=1$$

$$E(Y|X = x, S = 0, P(Z) = p) = \alpha_0 + X\beta_0 + \rho_0 \left(\frac{\Phi(\widehat{\mu_s Z})}{1-\Phi(\widehat{\mu_s Z})} \right) \text{ for } S=0;$$

4. From step 3 we obtain $\widehat{\alpha_1 + \varphi}, \widehat{\alpha_0}, \widehat{\beta_1}, \widehat{\beta_0}, \widehat{\rho_1}, \widehat{\rho_0}$ that permit to calculate the MTE. So, keeping the mean values of all variables in the X vector, the MTE is:

$$\widehat{MTE}(X = x, U_s = u_s) = (\widehat{\alpha_1 + \varphi} - \widehat{\alpha_0}) + \bar{X}(\widehat{\beta_1} - \widehat{\beta_0}) + (\widehat{\rho_1} - \widehat{\rho_0})\Phi^{-1}(u_s)$$

Where $U_s = 1 - \Phi(\widehat{\mu_s Z})$, is the probability not to be treated. The MTE is evaluated at the margin of indifference, therefore at the values of the unobservables that make the individual indifference between enrolment or not. $\Phi^{-1}(u_s) = -\widehat{\mu_s Z}$.

The coefficient $(\widehat{\rho_1} - \widehat{\rho_0})$ if statistically significant indicates that there is selection on unobservables. If $\widehat{\rho_1} < \widehat{\rho_0}$, there is selection on returns.

4.4 ESTIMATION RESULTS

Hereafter I present estimation results obtained applying the above stated model.

4.4.1 First-stage regression

Table 5. Probit regression (1)

Probit regression					Number of obs= 409	
					LR chi2(33)=	74.27
					Prob > chi2=	0.0001
Log likelihood = -237.88064					Pseudo R2=	0.135
schooling	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
<i>fees</i>	-0.0010449	0.0010166	-1.03	0.304	-0.0030374	0.0009476
<i>grants</i>	0.0003464	0.0003433	1.01	0.313	-0.0003264	0.0010193
<i>loans</i>	-0.0009504	0.0005929	-1.60	0.109	-0.0021124	0.0002117
<i>distance</i>	-0.0099131	0.009275	-1.07	0.285	-0.0280917	0.0082656
<i>sex</i>	-0.0057636	0.1424166	-0.04	0.968	-0.284895	0.2733677
<i>age</i>	0.0520573	0.2215278	0.23	0.814	-0.3821291	0.4862437
<i>qual</i>	0.1067409	0.2916708	0.37	0.714	-0.4649234	0.6784052
<i>jbstat</i>	0.2846678	0.1030772	2.76	0.006	0.0826402	0.4866955
<i>unat18</i>	-2.59579	1.319089	-1.97	0.049	-5.181157	-0.0104232
<i>_lmqfedhi_2</i>	-0.1402572	0.1626493	-0.86	0.389	-0.4590441	0.1785296
<i>_lmqfedhi_3</i>	-0.2764767	0.1868908	-1.48	0.139	-0.642776	0.0898225
<i>fisit</i>	0.0469359	0.1545225	0.30	0.761	-0.2559226	0.3497944
<i>_lincomecla_2</i>	0.6597746	0.6583042	1.00	0.316	-0.6304779	1.950027
<i>_lincomecla_3</i>	0.6359548	0.7071711	0.90	0.368	-0.7500751	2.021985
<i>_lcohort_2</i>	8.315	3.894383	2.14	0.033	0.6821491	1.594785
<i>_lcohort_3</i>	7.639767	3.507563	2.18	0.029	0.7650699	1.451446
<i>_lcohort_4</i>	4.042627	1.626069	2.49	0.013	0.8555899	7.229664
<i>_lcohort_5</i>	4.969295	2.19518	2.26	0.024	0.6668217	9.271768
<i>_lcohort_6</i>	1.97956	0.4996489	3.96	0.000	1.000267	2.958854
<i>_lcohort_7</i>	0.9706057	0.7305394	1.33	0.184	-0.4612253	2.402437
<i>_lcohort_8</i>	1.652889	1.179638	1.40	0.161	-0.6591595	3.964938
<i>_lcohort_9</i>	2.54744	1.558009	1.64	0.102	-0.506202	5.601082
<i>_lcohort_10</i>	-3.058033	1.469071	-2.08	0.037	-5.937359	-0.1787072
<i>_lregion2_2</i>	0.1835043	0.3688069	0.50	0.619	-0.539344	0.9063526
<i>_lregion2_3</i>	-0.1275673	0.3949873	-0.32	0.747	-0.9017282	0.6465936
<i>_lregion2_4</i>	-0.3590907	0.3842141	-0.93	0.350	-1.112137	0.3939551
<i>_lregion2_5</i>	-0.0727445	0.3958803	-0.18	0.854	-0.8486556	0.7031666
<i>_lregion2_6</i>	-0.053179	0.3783943	-0.14	0.888	-0.7948183	0.6884603
<i>_lregion2_7</i>	0.265451	0.3763639	0.71	0.481	-0.4722087	1.003111
<i>_lregion2_8</i>	-0.1911401	0.3516647	-0.54	0.587	-0.8803903	0.4981102
<i>_lregion2_9</i>	-0.0926924	0.3920143	-0.24	0.813	-0.8610264	0.6756415
<i>_lregion2_10</i>	0.1724351	0.3835363	0.45	0.653	-0.5792822	0.9241524
<i>_lregion2_11</i>	0.9221118	0.4160514	2.22	0.027	0.106666	1.737558
<i>_cons</i>	4.051678	2.181731	1.86	0.063	-2.244366	8.327792

From *Table 5*, the first aspect to be noticed is the statistical insignificance of the instrument chosen. This is a fundamental problem because under this condition the exclusion restriction is not satisfied. Therefore, the model is not identified since it is not possible to extrapolate any exogenous variation in the probability of being enrolled into university. Consequently, the endogeneity problem cannot be solved and this precludes the estimation of the Marginal Treatment Effect or other parameter of interest.

Anyway, the attempt is to try to analyse the evidence obtained to investigate, for what possible, the source of the problem.

Therefore, I try to specify alternative regression equations modifying the conditioning set, in order to see if the R-squared improves. In doing this I proceed without going against the theoretical setting that justifies the application of the model as presented in last chapters. Consequently, as seen in *Chapter 3*, familiar characteristics are typically controlled for in the estimation of the probability of enrolling into university, since they are a factor driving heterogeneity in enrolment among students.

Dropping *sex* and *age* (controls with the high level of the p-value) and *qual* (highest grades of secondary educational qualification achieved) does not lead to any improvement in terms of pseudo R-squared. The reason for dropping *qual* arises from the concern that it could be a “bad control”, meaning that it could be itself an outcome of another variable, e.g. *mqfedhi* (mother education) in this case. The doubt arises from empirical evidence presented in *Chapter 3*. The literature suggests that qualifications obtained at 16 years old have a great impact on university enrolment but at the same time, children with highest qualification are those belonging to higher social class’ families, with more propensity for education. It is also to be considered that this variable gives just a rough measure of prior educational attainment as explained in previous section. A more precise indicator would control for the number of subjects passed with the highest grades in order to have a proxy for the measure of ability at 16 years old.

The major improvement in terms of pseudo R-squared is achieved when the instrument *distance to nearest university* is dropped. The concern about the validity of this instrument is that from BHPS data it was not possible to retrieve the urban residence of each individual at the age of 14 or 16 (as for example considered in Carneiro, Heckman and Vytlačil, 2011). Distance has been measured considering local authority district of residence in the

“eligibility” year. This could mean that *distance* appears to be not exogenous if for example the family has decide to move after secondary school to permit continuation into further education of the children.

The alternative specification of the first-stage regression dropping *distance* has not lead to any substantial improvement in the statistical significance (or value of) of other coefficients. Number of observations also remain substantially the same (N=426).

Last consideration focuses on sample size. Keeping some variable in the conditioning set, like *mqfedhi* and *fisit*, shrinks the sample a lot (see *Table 6.- 7.*)

Table 6.

	<i>financial situation report by father</i>		
schooling	0 (bord/bad)	1 (wealthy/rel)	Total
0	68	105	173
1	87	184	271
Total	155	289	444

Table 7.

	highest educational qualification			
schooling	1(further/high)	2(sec or app.)	3(no qual)	Total
0	55	73	64	192
1	126	94	59	279
Total	181	167	123	471

Indeed, the Probit first-stage regression considerably change excluding these controls (see *Table 8.*) (the sample size almost double), maybe indicating the poor performance of the Maximum Likelihood estimator on small samples.

Table 8. Probit regression (2)

Probit regression					Number of obs= 806	
					LR chi2(27)=	64.75
					Prob > chi2=	0.0001
Log likelihood = -509.50087					Pseudo R2=	0.0597
schooling	Coef.	Std. Err.	z	P>z	[95% Conf.Interval]	
<i>fees</i>	-0.0010294	0.0005196	-1.98	0.048	-0.0020478	-0.000011
<i>grants</i>	0.0003303	0.0001946	1.70	0.090	-0.0000511	0.0007117
<i>loans</i>	-0.0006339	0.0003663	-1.73	0.084	-0.0013519	0.0000841
<i>jbstat</i>	0.2682615	0.0694446	3.86	0.000	0.1321525	0.4043705
<i>unat18</i>	-1.785101	0.7651616	-2.33	0.020	-3.28479	-0.2854122
<i>_lincomecla_2</i>	0.4007677	0.3603031	1.11	0.266	-0.3054134	1.106949
<i>_lincomecla_3</i>	0.5930517	0.4253723	1.39	0.163	-0.2406626	1.426766
<i>_lcohort_2</i>	5.576763	2.271392	2.46	0.014	1.124917	1.002861
<i>_lcohort_3</i>	5.052427	2.056147	2.46	0.014	1.022454	9.0824
<i>_lcohort_4</i>	2.498071	1.016143	2.46	0.014	0.506468	4.489675
<i>_lcohort_5</i>	3.45908	1.379122	2.51	0.012	0.756051	6.162109
<i>_lcohort_6</i>	1.021856	0.3470302	2.94	0.003	0.3416895	1.702023
<i>_lcohort_7</i>	0.8693246	0.381167	2.28	0.023	0.1222511	1.616398
<i>_lcohort_8</i>	1.354933	0.6796123	1.99	0.046	0.0229172	2.686949
<i>_lcohort_9</i>	1.981313	0.8985634	2.20	0.027	0.2201611	3.742465
<i>_lcohort_10</i>	-2.114388	0.8657716	-2.44	0.015	-3.811.269	-0.4175071
<i>_lregion2_2</i>	0.380157	0.2705667	1.41	0.160	-0.1501439	0.9104579
<i>_lregion2_3</i>	0.2486645	0.2884461	0.86	0.389	-0.3166795	0.8140085
<i>_lregion2_4</i>	0.0108054	0.2841473	0.04	0.970	-0.5461131	0.567724
<i>_lregion2_5</i>	0.2687559	0.2952604	0.91	0.363	-0.3099439	0.8474557
<i>_lregion2_6</i>	0.1007669	0.2906205	0.35	0.729	-0.4688389	0.6703726
<i>_lregion2_7</i>	0.4182252	0.2883784	1.45	0.147	-0.146986	0.9834364
<i>_lregion2_8</i>	0.2528414	0.263979	0.96	0.338	-0.2645479	0.7702307
<i>_lregion2_9</i>	0.0189919	0.2779702	0.07	0.946	-0.5258197	0.5638034
<i>_lregion2_10</i>	0.3553505	0.2868336	1.24	0.215	-0.206833	0.9175339
<i>_lregion2_11</i>	0.5693553	0.28003	2.03	0.042	0.0205066	1.118204
<i>_lregion2_12</i>	-0.3669044	0.3947927	-0.93	0.353	-1.140684	0.4068751
<i>_cons</i>	2.818853	1.25253	2.25	0.024	3.639397	5.273766

Estimation appears to be very sensitive to the increase of sample size (N=806) that has almost doubled. It results in a remarkable decrease in the p-values of the instruments at the 5% level, with fees becoming weakly significant. Dropping the dummy for *incomeclass* generate an even greater impact on estimates' precision. (see Table 9.)

Table 9. Probit regression (3)

Probit regression					Number of obs= 806	
					LR chi2(25)=	62.73
					Prob > chi2=	0.0000
Log likelihood = -510.50889					Pseudo R2=	0.0579
Schooling	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
Fees	-0.0006281	0.0002431	-2.58	0.010	-0.0011046	-0.0001516
Grants	0.0001259	0.0000712	1.77	0.077	-0.0000136	0.0002654
Loans	-0.0007222	0.0002429	-2.97	0.003	-0.0011983	-0.0002461
Jbstat	0.2725542	0.069329	3.93	0.000	0.1366718	0.4084366
unat18	-1.722009	0.4761413	-3.62	0.000	-2.655229	-0.7887896
_lcohort_2	5.426281	1.439074	3.77	0.000	2.605747	8.246815
_lcohort_3	4.955715	1.307045	3.79	0.000	2.393954	7.517477
_lcohort_4	2.61174	0.6860248	3.81	0.000	1.267156	3.956323
_lcohort_5	3.588679	0.9158136	3.92	0.000	1.793717	5.383641
_lcohort_6	1.25975	0.2749014	4.58	0.000	0.7209529	1.798547
_lcohort_7	0.7701347	0.277317	2.78	0.005	0.2266034	1.313666
_lcohort_8	1.215305	0.4299538	2.83	0.005	0.3726116	2.057999
_lcohort_9	1.861179	0.5578092	3.34	0.001	0.7678927	2.954465
_lcohort_10	-2.072444	0.5785431	-3.58	0.000	-3.206368	-0.9385202
_lregion2_2	0.3764103	0.2704252	1.39	0.164	-0.1536133	0.9064339
_lregion2_3	0.2286591	0.2879848	0.79	0.427	-0.3357808	0.7930989
_lregion2_4	0.0019163	0.2835529	0.01	0.995	-0.5538373	0.5576698
_lregion2_5	0.2640437	0.2950912	0.89	0.371	-0.3143243	0.8424118
_lregion2_6	0.0899581	0.2900213	0.31	0.756	-0.4784732	0.6583895
_lregion2_7	0.4034632	0.2882444	1.40	0.162	-0.1614853	0.9684118
_lregion2_8	0.2558676	0.2636728	0.97	0.332	-0.2609217	0.7726569
_lregion2_9	0.0147634	0.2772579	0.05	0.958	-0.528652	0.5581789
_lregion2_10	0.349902	0.2865304	1.22	0.222	-0.2116872	0.9114912
_lregion2_11	0.5567052	0.2796686	1.99	0.047	0.0085648	1.104846
_lregion2_12	-0.3908767	0.3943201	-0.99	0.322	-1.16373	0.3819765
_cons	2.764955	7.894069	3.50	0.000	1.217746	4.312164

Distance is not included in Table 9.; anyway running the Probit estimates with *Distance* confirms the evidence before presented. The effect of *Distance* is not statistically significant at 5% level (P-value 0.319) and its inclusion does not alter other coefficients' estimates.

In conclusion, estimates are very sensitive under different specification and sample size preventing a logical interpretation of results and above all, a valid continuation of the analysis. Indeed, what emerges from Table 5. is that not only there is no exclusion restriction

available, but also there is not a sufficient evidence in the data that allows to conclude that mother education, qualification obtained at 16 years old and the subjective consideration of the household's financial situation have an impact in university enrolment.

The only variables that throughout all specifications have an impact (respectively positive and negative on the enrolment probability) are the employment status of the individual in the eligible year (*jbstat*) and the unemployment rate in the same year for individuals aged 15-19 (*unatl8*). This probably indicates that students having the possibility of working before university have more probability of enrolling because they are able to gain some money that will help in facing the financial sacrifice (or acquiring more independence from parents). This interpretation would be in line with the negative impact of the unemployment rate.

In *Table 9*, even if the instruments acquire some significance (at least for fees and loans), this is not sufficient to prove their validity and that they provide the necessary variation over time to define a common support of the propensity score which is large enough to estimate the MTE at all point of the margin of indifference.

Assuming the instrument are valid, therefore fees, grants and loans are uncorrelated with unobservable components of the outcome equation, it may be that their relevance is small. By just looking at the correlation between the instruments and the endogenous regressor (schooling) it may raise the doubt of dealing with weak instruments. (*Table 10*.)

Table 10. Correlation coefficients

	schooling	fees	grants	loans
<i>schooling</i>	1.0000			
<i>fees</i>	-0.0005	1.0000		
<i>grants</i>	0.0131	-0.4595	1.0000	
<i>loans</i>	0.0248	0.5502	-0.4648	1.0000

First-stage regression leads to determine the common support of the propensity scores over which the margins of indifference are identified.

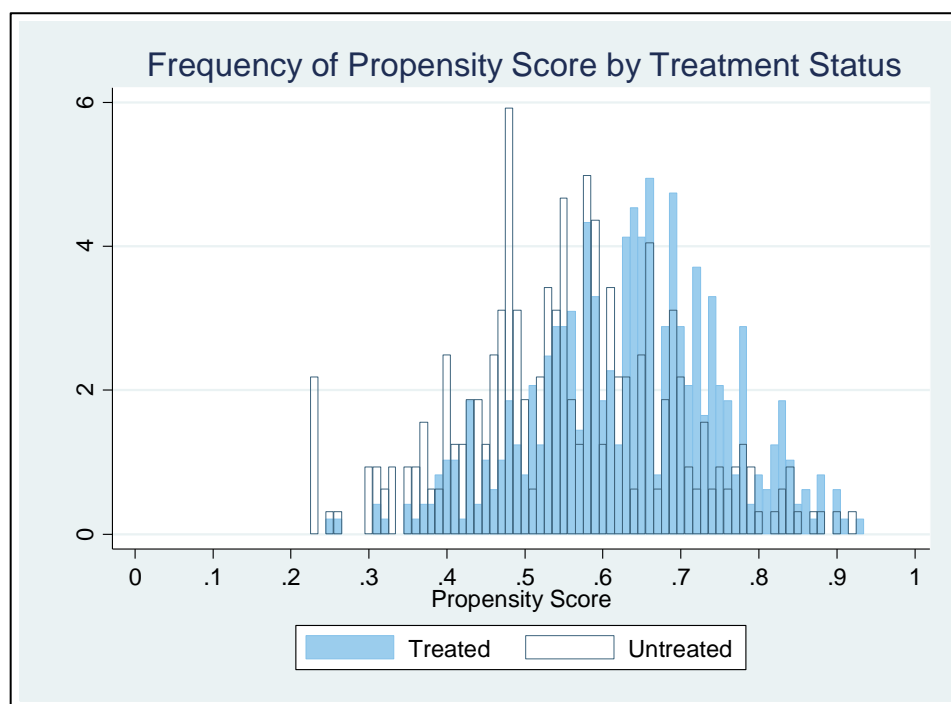
With the data available no margin can be identified in order to calculate the parameter of interest, but it is possible to see the graph of the common support to understand the implications of dealing with non-performing instruments. As described in *Chapter 2*., having multiple continuous instruments is a fundamental requirement in order to enlarge the common support of the propensity score. This allows to recover all parameters of interest, but in

particular to determine the treatment effect for all individuals selected who find themselves at different margins of indifference, identified by all values of the unobserved components that make them less likely to participate.

The larger the support, the more margins are identified and also the more precise the estimates are; indeed, sample size is also affected by the common support that we are able to define. For all ranges of the p-score in which we detect observations belonging to just one treatment status group, those observations have to be deleted. This can exert a negative impact especially when dealing with small sample as in the case of the current analysis.

Below is shown the *Graph 1*. of the common support of the propensity score resulting from the first-stage regression of *Table 9*. The common support in this case is in the range [0.35 – 0.8] and just in some isolated point outside it. Therefore, we are quite far from achieving a full support situation.

Graph 1. Common support of the propensity score



The selection of the right instrument is a great concern when Instrumental Variable, both local and linear in general, are applied. For instance, Carneiro and Heckman (2002) analyse the nature of credit constraints effect on the decision to enrol into post-secondary education distinguishing the effect of short-run liquidity constraint from long-term ones. In their analysis, they criticize the choice of common instruments in the literature of returns to

schooling, like tuition and distance to college arguing that these are invalid instruments. In particular, they develop a two-period model of credit-constraint schooling taking into consideration college quality in individuals' decision. They argue that the invalidity of instruments commonly chosen derives from the fact that these are correlated with school quality, which enters into the potential wage equation.

Obviously their analysis reflect the American situation, but it is not to exclude that this can a point of debate also in the case of the U.K.; in particular, after the reform of 2006/07 university are made free to establish the amount of fees they want respecting the maximum one fixed by the reform. Under this scheme, it would be reasonable to assume that highest quality college feel free to charge higher amounts justifying higher quality of teaching and if this is the case, tuition are related to university's quality. Even if the analysis here conducted involve access costs up to 2002/03, in the BHPS there are not any variable that indicate or proxy the quality of the institute attended, therefore it would not suitable to perform analysis for subsequent cohort under the framework developed here.

4.5 POSSIBLE EXTENSION AND THE EX-ANTE EVALUATION OF RETURNS TO EDUCATION

As described in *Chapter 2.*, the estimation of the Marginal Treatment Effect using a structural model like the Generalized Roy model, that allows to account for the decision process of the individual when they self-select into schooling, is a methodology that under the stated assumptions allow to perform a wide range of evaluations.

With the appropriate data indeed, it is possible not only to estimate the causal effect of a certain "treatment", which in this case is the possess of a Higher Educational qualification, but also try to identify the effect for a particular set of individual that stays on "the margin". Having the possibility to recognize this class of individuals permit develop some considerations about the mechanism to enlarge participation into higher education. This can be achieved through some policy reforms, that can induce individuals to enrol.

This is the point in which ex-post evaluation can be linked with an ex-ante approach. Unfortunately, the data used in the analysis presented in this dissertation have demonstrated to be not suitable for this kind of study. As a consequence, with a small sample and weak instruments it is not possible to perform an ex-ante evaluation of potential new policy rules concerning access costs to higher education.

If this would have been possible, some other parameters would have been considered, like the Policy Relevant Treatment Effect. By simulating a marginal change in policy affecting tuition, therefore in the instruments, it would have been possible to see: how the propensity of enrolling into higher education change, and what would have been the potential outcome change associated with new value of the propensity score.

The shortcoming of this approach is that all relies on the availability of good instruments, which are typically difficult to be found.

In the literature indeed, the largest part of studies that focus on the evaluation of policy reforms in the higher education sector in the U.K. are concentrated on the estimation of the causal effect of a particular set of access rule on the probability to enrol, without extending the analysis on the estimation of the returns to this “treatment”.

Analysing the effect on future wages of holding an higher educational qualification for those individual that has been induced into participation by the set of reforms in act at the time they were eligible, requires a great quantity (and quality) of data.

An alternative way of deriving an ex-ante evaluation framework, is that of taking inspiration from the existing studies that evaluate the effect of tuition’s reform on university enrolment to develop and validate a structural model that allows to perform an ex-ante analysis.

As seen in the first chapter, in the literature many empirical studies that are aimed to perform ex-ante evaluation start from a natural or quasi-natural experiment as the basis to validate a structural model.

For example, Attanasio, Meghir and Santiago (2012) develop a dynamic school participation model to study the effect of a monetary subsidy on school participation. Obviously, in that case they rely on a solid base in order to validate their formulation, the randomized social experiment Progreso. Anyway, we have seen that all kind of ex-post evaluation studies can be used as the ground against which benchmark a structural model, especially if a certain policy have been evaluated by different authors and with different methodologies.

This for example could be applied also for what concerns the estimation of the latest reforms happened in the U.K.; Dearden, Fitzsimons, Wyness (2014) have studied the effect of the re-introduction of grants for low-income students, taking high-income students as the

control group. The causal effect results to be positive and even if it refers only to the a particular measure in a particular time it would be a good basis to design a dynamic/discrete choice model that would be used to test ex-ante the effect of different potential values of grants.

Even if limited only to a particular measure, it would be a starting point to conduct ex-ante analysis when other kind of techniques are difficult to be applied.

CONCLUSIONS

This thesis has investigated the methods that allows to perform ex-post evaluation and to exploit the results obtained to conduct ex-ante evaluation.

The fundamental difference between ex-post and ex-ante approaches is that the latter is realized by modelling the behaviour and decision making process of agents. This is typically achieved through structural models, estimated parametrically or non-parametrically. Structural models describing behaviour of agents take the form of discrete choice models based on a latent variable specification. These models, in particular under parametric specifications, rely on some assumptions concerning functional forms and the distribution of unobservables that, even if strong, allow the parametric specification to be a powerful tool to predict the effects of policies never implemented.

A part from the way in which structural models are estimated, they need to satisfy an exclusion restriction that ensure identification. The design of a structural model involves an outcome equation that can be defined, for instance, as a utility function maximized under the budget constraint. The presence of the budget constraint, besides defining individual's decisional process, allows the identification. This is achieved thanks to the presence of a policy instrument that provides an exogenous source of variation, meaning that it affects only the budget constraint without directly influencing the outcome-specific equation. The source of variation can be also provided by an element different from the policy instrument object of the analysis, but able to provide a policy-relevant variation, therefore a variation that is isomorphic to that of the policy instrument.

In this dissertation returns to higher education in the U.K. has been evaluated. The motivation relies on the empirical evidence observed in this sector over last decades, starting from the 1960s. The Government has faced an intensive increase in the volume of students enrolled into undergraduate courses, with the consequence that resources addressed to funding them started to be not sufficient. The solution to the funding problems materialized into a series of reforms that from the 1990s onwards were issued. They introduced and subsequently modified access costs and benefits, like fees, grants and loans.

After the introduction of these reforms, several authors examined the causal effect exerted by new access rules on the enrolment rate, guided by the evidence suggested by data: the gap in the participation rate between “low-income” and “high-income” individuals. The literature

presented in previous chapters demonstrates that financial measures in favour of less-wealthy students are effective in reducing the existing gap and that familiar backgrounds seems to play a crucial role in children's future choice. Therefore, we are dealing with a sector in which the Government has to deal with multiple objectives, consisting in the promotion of participation among all social classes, and the attribution of the appropriate amount of resources to all institutes in order to promote an efficient service. This scenario raises questions related to what could be the right policy instrument that can ameliorate this situation.

Returns to higher education in the U.K. has been studied here applying the Generalized Roy model, estimated parametrically through the two-step estimator method. It implies that the "budget constraint" defined through a latent index structure, is estimated through Maximum Likelihood obtaining different values of the propensity scores; after this, the alternative-specific outcome equation is estimated through OLS including the appropriate correction term.

The Generalized Roy model has been used in the literature to study the returns to education under essential heterogeneity, a situation in which returns are assumed to vary between individual given their unobservables characteristics. This framework has several implications, since it may lead to biased estimates when applying traditional estimation methods like OLS or IV. Individuals that normally differ in unobservable characteristics, like ability, motivation, determination and other cognitive and non-cognitive skills are driven by these factors during their decisional process, creating the ability bias commonly known. These factors cause also the heterogeneity in returns, since they are normally affected by ability and other personal characteristics, besides education, experience and labour market's conditions. Therefore, if individuals are aware of their idiosyncratic returns to education, they can act on this knowledge at the time when they decide to enrol, generating another kind of bias: the selection on unobservables bias; if more clever or capable students are those who enrol, we say that there is selection in gains.

The existence of this mechanism has implications on the kind of parameter estimated. As discussed in *Chapter 2.*, under essential heterogeneity the only relevant parameter that can draw a picture of this situation is the marginal treatment effect, a parameter that define the returns to education for individuals at the margin of indifference, thus individuals which are indifferent between participating or not. The marginal treatment effect varies with respect to different values of the unobserved component that influence the probability of participating,

therefore it tells what would be the return for individuals with different unobservable characteristics if they would be induced to college by a marginal change in their probability to enrol.

Unfortunately, the attempt to estimate this parameter with the data chosen (the BHPS) has been inconclusive. The main problems faced in conducting this analysis concern in particular the data management. Even if data from eleven subsequent years have been selected, the relevant sample has revealed to be too small and this can be considered the first cause for non-significant results.

Moreover, the availability of necessary variables was also limited. Several variables had to be derived or even if available, reported a high number of missing values, involving the choice of alternative measures that serve as a proxy (e.g. qualification achieved at 16 years old); nevertheless, in the case of the instrumental variable *distance*, the lack of postcodes identifying the area of residence, generated difficulties in the calculation of the distance from the nearest institute.

The analysis therefore stopped at the first stage regression, from which the only conclusion is that there is not a sufficient evidence of statistical significance. This implies that the model is not identified, since the instruments seems not to provide the necessary exogenous variation to satisfy the exclusion restriction.

A potential replication of this analysis should, first of all, rely on another dataset, that has to be panel in order to observe an individual from 18 to $18+t$ years old, age in which labour wages are observed, and has to provide the necessary variables to create subsamples of treated and untreated of a suitable size.

Due to these shortcomings, the analysis has stopped, and the ex- ante evaluation could not be performed. As presented in the theoretical framework described in this thesis, the identification of the marginal treatment effect in general allows to perform not only ex-post evaluation, but also the ex-ante one. Indeed, it is possible to estimate how the probability to enrol would change under marginal change in the value of the policy instruments. Given new potential value of the propensity score, we can define the mean effect for an individual that is induced to participate by a certain intervention, as expressed by the policy relevant treatment effect.

In any case, if necessary data to analyse the returns to education were not available, it is still possible to develop a useful ex-ante evaluation, in order to analyse, for instance, the effect of certain reforms on the participation rate among the “low-income” individuals, being them the category still less present in the higher education sector. Exploiting methodologies explained in *Chapter 1.*, it is not to exclude that a valid structural model can be selected, validated (relying on a reasonable number of empirical studies in the literature) and used to run counterfactual analysis.

REFERENCES

1. Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal*, 114(494), C52-C83.
2. Angrist, J., and Imbens, G. (1995). Identification and estimation of local average treatment effects.
3. Attanasio, O. P., Meghir, C., and Santiago, A. (2012). Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progreso. *The Review of Economic Studies*, 79(1), 37-66.
4. Barr, N. A., and Crawford, I. (2005). *Financing higher education: Answers from the UK*. London: Routledge.
5. Becker, G. S. (1967). *Human capital and the personal distribution of income: An analytical approach* (No. 1). Institute of Public Administration.
6. Bernal, R. and Keane, M. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1), pp.164-189.
7. Björklund, A., and Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, 69(1), 42-49.
8. Blundell, R. (2006). Earned income tax credit policies: Impact and optimality. *Labour Economics*, 13(4), pp.423-443.
9. Blundell, R., and Dias, M. C. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3), 565-640.
10. Blundell, R., Bozio, A. and Laroque, G. (2013). Extensive and Intensive Margins of Labour Supply: Work and Working Hours in the US, the UK and France. *Fiscal Studies*, 34(1), pp.1-29.
11. Bolton, P. (2015). Tuition fee statistics.

12. Bourguignon, F., and Ferreira, F. (2003). Ex ante evaluation of policy reforms using behavioral models. *The impact of economic policies on poverty and income distribution*, 123-142.
13. Brewer, M., Duncan, A., Shephard, A. and Suárez, M. (2006). Did working families' tax credit work? The impact of in-work support on labour supply in Great Britain. *Labour Economics*, 13(6), pp.699-720.
14. Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801-1863.
15. Card, D. (2001). *Estimating the return to schooling: Progress on some persistent econometric problems* (No. w7769). National Bureau of Economic Research.
16. Carneiro, P. (2003). Heterogeneity in the returns to schooling: implications for policy evaluation. (Doctoral dissertation, University of Chicago, Department of Economics August 2003).
17. Carneiro, P., Heckman, J. J., and Vytlacil, E. (2001). Estimating the return to education when it varies among individuals. *University of Chicago*, working paper.
18. Carneiro, P., and Heckman, J. J. (2002). The Evidence on Credit Constraints in Post-Secondary Schooling. *The Economic Journal*, 112(482), 705-734.
19. Carneiro, P., Heckman, J. J., and Vytlacil, E. (2003). Understanding what instrumental variables estimate: Estimating marginal and average returns to education. *processed, University of Chicago, The American Bar Foundation and Stanford University*, 19.
20. Carneiro, P., Heckman, J. J., and Vytlacil, E. (2010). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica*, 78(1), pp.377-394.
21. Carneiro, P., Heckman, J. and Vytlacil, E. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6), pp.2754-2781.
22. Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 431-457.

23. Corver, M. (2010). Trends in young participation in higher education: core results for England. *HEFCE publication*, 3, 1-36.
24. Crawford, C., and Greaves, E. (2015). Socio-economic, ethnic and gender differences in HE participation.
25. Dearden, L., Fitzsimons, E., and Wyness, G. (2011). The impact of tuition fees and support on university participation in the UK (No. W11/17). *IFS Working Papers*.
26. Dearden, L., Fitzsimons, E., and Wyness, G. (2014). Money for nothing: Estimating the impact of student aid on participation in higher education. *Economics of Education Review*, 43, 66-78.
27. Di Nardo, J., and Lee, D. S. (2011). Program evaluation and research designs. *Handbook of labor economics*, 4, 463-536.
28. Di Porto, E., Elia, L., and Tealdi, C. (2013). The “emersion” effect: an ex post and ex ante social program evaluation on labor tax evasion in Italy.
29. Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *The American Economic Review*, 1241-1278.
30. Geyer, J., Hann, P., and Wrohlich, K. (2012). Labor supply of mothers with young children: Validating a structural model using a natural experimental. In *Conference Proceedings of IZA Workshop: Recent Advances in Labor Supply Modeling, Institute for the Study of Labor (IZA), Bonn*.
31. Great Britain. Department for Business Innovation, Department for Business, and Great Britain. Parliament. (2011). *Higher Education: Students at the Heart of the System* (Vol. 8122). The Stationery Office.
32. Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
33. Greene, W. H. (2005). Censored data and truncated distributions. *Available at SSRN* 825845.
34. Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, 1-22.
35. Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

36. Heckman, J. J. (2001). Econometrics Counterfactuals and Causal Models.
37. Heckman, J. J. (2010). *Building bridges between structural and program evaluation approaches to evaluating policy* (No. w16110). National Bureau of Economic Research.
38. Heckman, J. J., and Honoré, B. E. (1990). The empirical content of the Roy model. *Econometrica: Journal of the Econometric Society*, 1121-1149.
39. Heckman, J., and Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and statistics*, 86(1), 30-57.
40. Heckman, J. J., and Robb, R. (1986). *Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes* (pp. 63-107). Springer New York.
41. Heckman, J. J., and Vytlačil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, 96(8), 4730-4734.
42. Heckman, J. J., and Vytlačil, E. (2000). Causal parameters, structural equations, treatment effects and randomized evaluations of social programs. *manuscript, University of Chicago*.
43. Heckman, J. and Vytlačil, E. (2001). Policy-Relevant Treatment Effects. *American Economic Review*, 91(2), pp.107-111.
44. Heckman, J. J., and Vytlačil, E. (2001b). Policy-relevant treatment effects. *The American Economic Review*, 91(2), 107-111.
45. Heckman, J. J., and Vytlačil, E. (2005). *Structural equations, treatment effects and econometric policy evaluation* (No. w11259). National Bureau of Economic Research.
46. Heckman, J. J., and Vytlačil, E. J. (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, 6, 4779-4874.
47. Heckman, J. J., and Vytlačil, E. J. (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative

- econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6, 4875-5143.
48. Heckman, J. J., Smith, J., and Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4), 487-535.
 49. Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Estimation of treatment effects under essential heterogeneity. *Retrieved October, 12, 2010*.
 50. Heckman, J., Tobias, J. L., and Vytlacil, E. (2001). Four parameters of interest in the evaluation of social programs. *Southern Economic Journal*, 211-223.
 51. Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945-960.
 52. House of common, (2012). Government reform of Higher Education: Government Response to the committee's 12th Report of Session 2010–12.
 53. Ichimura, H., & Taber, C. R. (2000). Direct estimation of policy impacts.
 54. Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), p.467.
 55. Keane, M. P., Todd, P. E., and Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. *Handbook of Labor Economics*, 4, 331-461.
 56. Marshall, A. (1890). *Principles of Economics*. Macmillan: London.
 57. Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings* (pp. 41-63). NBER.
 58. Pronzato, C. (2012). Comparing Quasi-Experimental Designs and Structural Models for Policy Evaluation: The Case of a Reform of Lone Parental Welfare. *IZA Discussion Papers 6803*, Institute for the Study of Labor (IZA).
 59. Ranjeeta, T. (2012). Conditional cash transfers to improve education and health: An ex ante evaluation of red de protección social, Nicaragua. *Health economics*, 1136-1154.

60. Thoresen, T. O., and Vattø, T. E. (2013). Validation of structural labor supply model by the elasticity of taxable income. *Statistics Norway Research Department Discussion Papers*, (738).
61. Todd, P. E., and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *The American economic review*, 1384-1417.
62. Todd, P. E., and Wolpin, K. I. (2010a). Ex Ante Evaluation Of Social Programs. *Annales d'économie et de statistique* , 91: 251–286.
63. Todd, P. E., and Wolpin, K. I. (2010b). Structural Estimation And Policy Evaluation In Developing Countries. *Annual Review Of Economics*, 2 (1): 21-50.
64. UCAS Analysis and Research. (2012). How have applications for full-time undergraduate higher education in the UK changed in 2012?
65. Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources*, 127-169.
66. Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1), 331-341.
67. Wasmer, E. (2012). Ex-ante and ex-post evaluation of the 1989 French welfare reform using a natural experiment: the 1908 social laws in Alsace-Moselle. In *2012 Meeting Papers* (No. 478). Society for Economic Dynamics.
68. Wyness, G. (2010). Policy changes in UK higher education funding, 1963-2009 (No. 10-15). Department of Quantitative Social Science-UCL Institute of Education, University College London.

WEBSITE CONSULTED:

- <http://ec.europa.eu/eurostat/data/database>
- <https://discover.ukdataservice.ac.uk/catalogue?sn=5151/>
- <https://www.gov.uk/>
- <https://www.hesa.ac.uk/>
- <https://www.iser.essex.ac.uk/bhps/>